Export markets and labor allocation in a low-income country

Brian McCaig and Nina Pavcnik

Online Appendix

Appendix A: Supplemental Tables for Sections III-IV

Appendix Table A.1: Growth of Vietnamese exports and BTA tariff changes

Dependent variable: Change in ln exports

	(1)	(2)	(3)	(4)
Industries	Traded	Manufacturing	Traded	Manufacturing
Destination market	US	US	EU13	EU13
Panel A: Change in ln exports 2001 to 2004				
BTA tariff change	-5.677***	-4.331*	0.372	0.142
	(1.474)	(2.111)	(0.675)	(1.070)
Observations	24	19	24	19
R-squared	0.283	0.119	0.009	0.001
Panel B: Change in ln exports 1997 to 2000				
BTA tariff change	-0.808	0.181	0.362	0.823
	(1.896)	(1.722)	(0.599)	(0.904)
Observations	24	19	24	19
R-squared	0.011	0.001	0.011	0.035

Notes: Robust standard errors in parentheses. ***, **, and * denotes significance at 1, 5, and 10 percent level, respectively. We use data on imports from Vietnam as reported by the U.S. and EU13 (EU15 excluding Belgium and Luxembourg for which data was not consistently available) in UNComtrade. We exclude industries for which imports were 0 for any of the years.

Appendix Table A.2: Descriptive statistics

	All		Pre BTA Round		Post BTA Round	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Self-employed	0.686	0.464	0.701	0.458	0.672	0.469
Worked in a household business	0.830	0.375	0.847	0.360	0.814	0.389
Indicator for urban	0.239	0.427	0.240	0.427	0.238	0.426
Age	37.8	11.1	37.4	11.0	38.3	11.1
Indicator for female	0.505	0.500	0.507	0.500	0.503	0.500
Indicator for ethnic minority	0.123	0.328	0.121	0.326	0.124	0.329
Indicator for completed primary education	0.288	0.453	0.297	0.457	0.280	0.449
Indicator for completed lower secondary education	0.300	0.458	0.292	0.455	0.307	0.461
Indicator for completed upper secondary education	0.199	0.399	0.185	0.388	0.212	0.409
Indicator for agriculture, forestry and aquaculture	0.542	0.498	0.561	0.496	0.524	0.499
Indicator for manufacturing	0.123	0.329	0.118	0.322	0.128	0.334
Indicator for services	0.327	0.469	0.313	0.464	0.341	0.474
Indicator for less than median distance from seaport	0.561	0.496	0.560	0.496	0.562	0.496
Hours per year (primary job)	1701	814	1730	792	1673	833
Indicator for more than one job					0.425	0.494
ln(hourly compensation)	1.368	0.726	1.234	0.791	1.494	0.634
Number of observations	248,795		152,388		96,407	

Notes: The sample consists of all workers from the 2002 and 2004 VHLSSs that worked and were 20 to 64 years of age inclusive at the time of the survey. The 2002 VHLSS is the pre BTA round and the 2004 VLHSS is the post BTA round. The number of observations for wages are lower: 46,309 and 29,758 in the 2002 and 2004 VHLSSs respectively. The total number of observations is slightly higher than in our regression results due to a small number of worker observations, 4, for which ethnicity data is missing. These observations are subsequently dropped from the regression analysis.

Appendix Table A.3: Differences in job characteristics between workers in household businesses and enterprises

	Traded		All		Manufacturing	
	Household businesses	Enterprises	Household businesses	Enterprises	Household businesses	Enterprises
Share self-employed	0.868	0.000	0.827	0.000	0.609	0.000
Mean annual hours worked	1,531	2,193	1,667	2,076	1,889	2,293
Mean highest grade completed	6.4	9.5	6.7	10.4	7.7	9.6
Share with more than one job	0.496	0.200	0.463	0.261	0.492	0.157
Mean annual total compensation	4,237	11,454	5,248	11,518	6,380	11,608
Share of wage workers that report receiving payments for …						
Holidays	0.150	0.846	0.188	0.861	0.296	0.863
Social insurance	0.001	0.034	0.002	0.037	0.003	0.032
Business trips	0.001	0.073	0.003	0.149	0.001	0.040
Other	0.117	0.585	0.156	0.605	0.168	0.606
Any reason	0.215	0.895	0.273	0.914	0.362	0.913

Notes: The sample is all workers age 20 to 64 in the 2002 VHLSS, with the exception of the share of workers with more than one job, which is based on workers age 20 to 64 in the 2004 VHLSS. The values are estimated using sampling weights. Social insurance payments mean payments received by workers, for example, for workplace injuries, not necessarily social insurance coverage. Compensation is reported in 000s of Vietnamese dong in January 2002 prices.

Appendix Table A.4: Additional Results for Employment in Household Businesses and Tariffs

Dependent variable: Indicator for working in a household business

	(1) Traded	(2) All	(3) Manufacturing
Panel A: Worker controls are omitted			
Tariff	0.212***	0.126***	0.162***
	(0.0140)	(0.0342)	(0.0210)
Observations	176,544	248,791	27,072
R-squared	0.404	0.580	0.245
Panel B: Remove mining of uranium and thorium ores from sample			
Tariff	0.209***	0.128***	0.156***
	(0.0144)	(0.0323)	(0.0197)
Observations	176,536	248,783	27,072
R-squared	0.419	0.594	0.299
Panel C: Pre-existing trends included			
Tariff	0.210***	0.138***	0.160***
	(0.0141)	(0.0271)	(0.0177)
(Change in industry ln employment)	-0.0160**	-0.0266*	0.0100
X (2004 indicator)	(0.00736)	(0.0133)	(0.0161)
(Change in industry self-employment rate)	0.00196	0.0799	0.0124
X (2004 indicator)	(0.0554)	(0.0677)	(0.0595)
(Change in industry mean grade completed)	0.0117**	0.0115*	0.0163*
X (2004 indicator)	(0.00488)	(0.00601)	(0.00864)
Observations	176,453	248,353	26,981
R-squared	0.418	0.592	0.298

Notes: Standard errors are clustered at industry level; ***, **, and * denotes significance at 1, 5, and 10 percent level, respectively. The sample is restricted to workers between the ages of 20 and 64 inclusive at the time of the survey. Column (1) includes all traded industries, column (2) includes all industries, and column (3) includes all traded manufacturing industries. All regressions include a rural indicator, industry, province, and year fixed effects, and regressions in Panels B and C also worker characteristics (age, age squared, education level indicators, female indicator, and ethnic minority indicator). The pre-existing trends in Panel C are calculated between 1993 and 1998 using the respective Vietnam Living Standards Surveys. The decrease in the number of observations in Panel C is due to a few small industries with missing information for pre-existing trends.

Appendix Table A.5: Hours Worked in Household Businesses and Tariffs

Dependent variable: Share of total industry hours worked within the household business sector

	(1) Traded	(2) All	(3) Manufacturing
Tariff	0.202***	0.119***	0.157***
	(0.0186)	(0.0379)	(0.0302)
Observations	68	120	44
R-squared	0.565	0.350	0.814
Industry fixed effects?	Yes	Yes	Yes
Year fixed effect?	Yes	Yes	Yes

Notes: Standard errors are clustered at industry level; ***, **, and * denotes significance at 1, 5, and 10 percent level, respectively. Column (1) includes all traded industries, column (2) includes all industries, and column (3) includes all traded manufacturing industries. Hours worked in the household business and enterprise sectors are estimated based on workers at 20 to 64 inclusive at the time of the 2002 and 2004 household surveys.

Appendix Table A.6: Industry Employment and Tariffs in Urban Areas

Dependent variable: Share of industry employment in the indicated set of industries

	(1) Traded	(2) All	(3) Manufacturing
Tariff	-0.00174	0.00163	0.00666
	(0.0136)	(0.00437)	(0.0421)
Observations	68	120	44
Within R-squared	0.001	0.006	0.003

Notes: Standard errors are clustered at industry level; ***, **, and * denotes significance at 1, 5, and 10 percent level, respectively. The dependent variable is the share of workers and is calculated as the number of urban workers in industry j divided by the total number of urban workers in the respective group. The total number of urban workers includes workers in (i) traded industries for column (1), (ii) all industries for column (2), and (iii) traded manufacturing industries for column (3). The industry employment shares are based on urban workers between the ages of 20 and 64 inclusive, calculated from the VHLSSs. All regressions include year fixed effects and industry fixed effects, using the within transformation.

Appendix B: Supplemental Material for Section V

This appendix provides a detailed description of the data and calculations used in Section V to estimate the aggregate labor productivity change in manufacturing associated with the reallocation of labor from household businesses to enterprises in response to the BTA.

B.1. Labor productivity gap between the household business and enterprise sectors

B.1.1. Basic Calculations

We compute the average revenue product of labor based on revenue per worker from firm-level data that covers all registered firms in the enterprise sector (the Vietnamese Enterprise Survey), and on aggregate revenue and the total number of workers in the household business sector from the household business and labor modules of the VHLSS. Most of our analysis relies on the 2001 Vietnamese Enterprise survey and the 2002 VHLSS.[1]

We calculate average revenue product of labor in the household business sector by summing annual revenue from all businesses reported in the household business module and dividing by the total number of workers in household businesses as reported in the labor module of the 2002 VHLSS.[2,3] Our estimate of revenue per worker in the household business sector is likely an overestimate because the business module in the 2002 VHLSS did not distinguish between

[1] The VHLSS has has a 12-month recall period and covers 2001/2002

[2] All revenue values are expressed in January 2002 Dong. For the VHLSS data, household business revenue is converted to January 2002 prices using monthly CPI data based on the month of interview. Information on the number of workers is based on employment in workers' primary job (i.e., employment based on their reported most time-consuming job). See discussion of employment in secondary jobs in Section B.1.5. We rely on the labor force module for employment because the 2002 household business module did not collect information about the number of workers in each business.

[3] Both estimates are weighted using survey sampling weights.

private enterprises (i.e., belonging in the enterprise sector) and household businesses. Its scope was *all* businesses run by households. This lowers the estimate of the gap in average labor productivity between the household business and enterprise sectors and thus leads to a cautious estimate of the gains from reallocation. See Section B.1.5 for further discussion.

We calculate revenue per worker in the enterprise sector based on total annual revenue divided by total employment in each enterprise at year-end using data covering 2001 from the 2002 Vietnam Enterprise Survey. Aggregate revenue per worker within the enterprise sector is the employment-weighted average of revenue per worker over all enterprises.

We use the labor force modules of the 2002 VHLSS to compute the wage gap between the enterprise and household business sectors. We compute the wage ratio using total earnings, which includes cash and in-kind wage/salary payments as well as other payments such as public holiday payments and social allowance payments, among wage workers in the two sectors.

These labor productivity gaps are reported in row 1 of Table 8 in the main text.

B.1.2. Existing evidence on the labor productivity gap

How do our estimates compare to those from the literature? Nationally representative data on household businesses (in other contexts also called informal firms or microenterprises) is rare, so we cannot compare our estimates to many previous studies. That said, our estimate is consistent with productivity gaps between informal and formal firms in other developing countries. We compare existing estimates to our unadjusted estimates in row one of Table 8 in the main text because the existing estimates do not adjust for worker heterogeneity across sectors. For example, Nataraj (2011) finds that output per worker is 12.4 times higher in formal firms than informal firms in India's manufacturing sector. La

Porta and Shleifer (2014) report gaps in value added per worker between formal and informal firms using firm surveys from 25 developing countries. They rely on World Bank Enterprise Surveys, which are not nationally representative and often exclude firms without any hired workers (i.e., firms that the literature often refers to as own-account workers). Consequently, the informal firms are larger on average, employing four workers, than in our context where the average number of workers is about 2. A big advantage of our data is that it is based on a census of all registered firms and a survey of household businesses based on a nationally representative household survey. With these caveats about the World Bank Enterprise Survey data in mind, value added per worker in formal firms is 6.7 times higher than in informal firms in the median country in their sample. This is lower than in our data, but the range across countries is large, from 5.3 in Mali to 14.3 in Tanzania, the 25th and 75th percentile countries, respectively. Our estimate of the productivity gap of a factor of 9 is within the range of estimates from other countries, comparable to their estimates for Angola and Kenya.

B.1.3. Adjusting the labor productivity gap for worker heterogeneity

The productivity gap in row 1 of Table 8 could reflect worker heterogeneity across the two sectors. We use micro-survey data on individuals from the labor force and education modules of the VHLSS to adjust this productivity gap for heterogeneity in hours worked and human capital between the household business and the enterprise sectors.

To begin with, the estimates reported in row 1 of Table 8 are based on revenue per worker rather than revenue per hour worked. However, enterprise sector workers work significantly more hours in the year than household business workers, 2,267 versus 1,825.[4] The difference in hours will reduce the productivity

[4] This is based on hours worked in a primary job and includes all workers regardless of age.

gap. We adjust the productivity measures from row 1 so that they reflect revenue per hour worked in each sector.

Second, we also adjust the productivity gap for differences in human capital across sectors. We follow Gollin et al. (2014), who adjust for the differences in average years of education across the agriculture and non-agriculture sectors. Specifically, let average human capital within a sector be given by $\exp(rsch_s)$ where r is the rate of return to a year of schooling and sch_s is the average years of education in sector s. We use a rate of return of 10 percent as in Gollin et al. (2014), based on the observation in Banerjee and Duflo (2005) that returns to schooling are estimated to be around 10 percent in most countries and don't vary much between low- and high-income countries. Using this approach, human capital in the enterprise sector is 1.2 times higher than in the household business sector.

The productivity gap with both of these adjustments is reported in row 2 of Table 8 and discussed in the main text. These adjustments matter. As noted there, worker heterogeneity accounts for 37 percent of the original ARPL gap and for 70 percent of the original wage gap.

One potential objection to the above approach to controlling for worker heterogeneity is that it only controls for two dimensions of worker heterogeneity. Alternatively, one can estimate the wage gap for working in the enterprise sector with Mincerian regressions, while controlling for worker heterogeneity in other dimensions, including location, gender, age, ethnic minority status, and industry affiliation. Vollrath (2014) and Herrendorf and Schoellman (2015) apply this approach to the wage gap between non-agriculture and agriculture. We estimate the following regression:

$$\ln w_i = \alpha + \beta_e enterprise_i + \theta X_i + \varepsilon_i$$

where w_i is the real hourly wage of worker i, $enterprise_i$ is an indicator variable for working in the enterprise sector, and X_i is a vector of worker characteristics, including education, location, gender, age, age squared, minority status, and industry affiliation. The wage gap for working in the enterprise sector (relative to the household business sector) is given by $\exp(\beta_e)$, where β_e is the coefficient on the indicator that a worker works in the enterprise sector.

We estimate the above Mincerian regression using data on manufacturing workers from the 2002 VHLSS. The estimates of the coefficient on the enterprise indicator β_e are reported in Appendix Table B.1. After controlling for additional dimensions of worker heterogeneity, the wage gap is a similar order of magnitude as the wage gap reported in Table 8. The estimates that are the most comparable to the wage ratio adjusted for human capital in row 2 of Table 8 are in column 2.[5] The estimates in the table are of a similar order of magnitude as the estimates based on the wage ratio adjusted for worker characteristics.[6] For example, the most conservative estimate of the wage difference, reported in column 6, which controls for education, location, demographics, and industry affiliation, suggests that workers in the enterprise sector earn 25 percent more per hour than observationally equivalent workers in the household business sector. Thus, estimates of the wage gap based on the Mincerian regressions, which simultaneously control for several dimensions of worker heterogeneity, yield similar findings as the approach taken in the text.

[5] Comparisons between estimates in column 2, which control for heterogeneity in education with years of education, and column 3, which control for heterogeneity in education with education indicators, suggest that these two approaches yield similar coefficients on the enterprise sector indicator. We therefore control for heterogeneity in education with education indicators, as we did in Section IV of the paper.

[6] The difference in the estimates between the wage gap and Mincerian-based estimates could be attributed to additional controls for observable worker characteristics, differences in functional form assumption, and the fact that the wage bill ratio is weighted by hours worked, while the Mincerian regression weights each hourly wage observation equally.

We also use the individual panel data to estimate the impact on wages while controlling for unobserved worker heterogeneity, which might be correlated with worker wages and sector of employment. The results are reported in Appendix Table B.2. Our sample is all individuals that worked for wages in manufacturing in *both* 2001/02 and 2003/04. Hence, the number of workers included is lower than in Appendix Table B.1. The estimated coefficients on an indicator for working in the enterprise sector from a regression that controls for unobserved worker heterogeneity with individual fixed effects are reported in columns 1 and 2. The wage gap persists, albeit it is smaller in magnitude. The estimates of the coefficient β_e suggest that wage workers that move between these two sectors earn about 9 percent more per hour when working in the enterprise sector. As the sample has changed relative to Appendix Table B.1, we also estimate β_e using the cross-sectional specification that uses the same set of worker covariates as column 6 in Appendix Table B.1 with the panel sample. The results in column 3 are based on both 2002 and 2004 data, while column 4 uses only 2002 data. The estimates of β_e are very similar to those using the cross section in column 6 of Appendix Table B.1 (0.202 and 0.191 versus 0.221). Hence, the change in sample is not the primary reason for the lower coefficient on working in the enterprise sector reported in columns 1 and 2. Instead, controlling for unobserved individual heterogeneity is the important driver.

The above analysis excludes self-employed individuals, as they do not work for wages. Hence, as an additional check relative to the wage regressions, we also include the self-employed in the above analysis by focusing on worker hourly income as a dependent variable. For wage workers, this is wage earnings as above. For the self-employed, we use self-reported profits. Specifically, we focus on manufacturing household businesses run by the manager as their primary job. Additionally, we restrict the sample of businesses to those for which no other

household member reports being self-employed in the same industry. This allows us to assign the reported profits to the manager without having to make assumptions on how to assign profits across multiple household members. In this analysis, we are treating the profits from the business as the manager's wage earnings.

These results are reported in Appendix Table B.3, which follows the same specifications as Appendix Table B.1. Interestingly, while the findings are in general similar to the findings with wages, the earnings gap is smaller in magnitude, at 11 percentage points, relative to the differences in wages only. This suggests that the managers of household businesses earn higher profits than observationally equivalent wage workers.

In summary, the above discussion provides the details of how we adjust the labor productivity gaps for worker heterogeneity. In the main text, we use these insights to emphasize that worker heterogeneity accounts for an important component of the ARPL and wage gaps across sectors, 37 percent and 75 percent, respectively. For the wage gap, worker heterogeneity plays an important role when we estimate the gap with differences in mean earnings between sectors, analogous to the ARPL gap, and when we estimate the gap using Mincerian wage or earnings regressions. Overall, this analysis highlights the importance of relying on information from micro-survey data to account for worker heterogeneity.

B.1.4. Measurement concerns

We address several concerns about measurement error. To begin with, one may be concerned about the measurement error associated with combining two different data sources.[7] For example, the surveys might measure revenue or

[7] See Nataraj (2011) and Hsieh and Olken (2014) for such an approach using Indian data and Ulyssea (2017) in Brazil.

earnings differently, affecting the estimates of the productivity gap. We perform several checks to address this potential concern.

First, we compare annual earnings per manufacturing worker (i.e., the numerator of the wage gap) in the enterprise sector across the two data sources, the 2002 VHLSS and the 2001 enterprise data. Mean annual earnings per worker were 11.6 million VND in the 2002 VHLSS as compared to 12.0 million VND in the 2002 Enterprise Survey, covering 2001. The similarity of these estimates suggests that workers in the 2002 VHLSS are reporting labor earnings consistent with the reports of labor expenses from firms in the 2002 Enterprise Survey. Thus, the wage gap estimates between the enterprise and household business sector are very similar if we use wage data for enterprises based on the VHLSS or the enterprise survey.

Second, the business module of a later VHLSS survey, the 2006 VHLSS, collected information on whether the business owned by a household was a household business or registered as an enterprise. [8] We use this data to estimate the ARPL gap between household businesses and private enterprises using this one survey. In the 2006 VHLSS, 1.5 percent of manufacturing businesses were private enterprises, emphasizing the prevalence of household businesses in manufacturing. Note that this calculation does not capture the productivity gap between household businesses and other types of firms in the enterprise sector. Specifically, it excludes state-owned enterprises, foreign-invested firms, and collective firms. These results are reported in column 1 of Appendix Table B.4. ARPL is 3.6 times higher among private, domestic enterprises than household businesses. The ratio drops to 2.8 after we adjust it for worker heterogeneity. The gap is slightly lower than in Table 8 since foreign-invested, state-owned, and collective enterprises are not included in the estimate of average productivity in the enterprise sector. This gap is computed with data from 4 years after the 2002

[8] The 2006 VHLSS has a 12-month recall period and was conducted from May through November.

VHLSS baseline year. For comparison purposes, we estimate the ARPL gap between the entire enterprise sector and household business sector for 2005/06, using the same procedure from Section B.1.1 and the same two data sources for 2005/06 as those used for 2002 gaps reported in Table 8. These estimates are reported in columns 2 through 4 of Appendix Table B.4. For example, the estimates of the ARPL gap for manufacturing as a whole in column 2 suggest that the ARPL gap between sectors is 5.0, 3.5 adjusted for worker heterogeneity. As expected, they are higher than the estimates that rely on private domestic enterprises alone. In addition, these estimates suggest that the ARPL gap has partially closed between 2001/02 and 2005/06.

Overall, the above discussion suggests that the productivity gap between workers in the enterprise and household business sectors is not likely driven entirely by differences in how firms report revenues in the VHLSS and the enterprise survey.

An alternative measurement concern is that very few microenterprise operators keep formal accounts and thus measurement error is more likely to affect our estimate of ARPL in the household business sector than in the enterprise sector. We discuss the potential measurement error in revenue and labor supply and consider adjustments to the ARPL gap to take these concerns into account. The results are reported in column 1 of Appendix Table B.5. The first row of the table reports the ARPL gap adjusted for worker heterogeneity that was discussed in Appendix B.1.3 and that is also reported in row 2, column 1 of Table 8. This estimate serves as the upper bound on the productivity gap.

First, de Mel, McKenzie and Woodruff (2009) provide experimental evidence on possible measurement error in revenues of microenterprises from Sri Lanka, suggesting that reported revenue in these businesses may underestimate true revenue by as much as 30 percent. Their sample covers microenterprises without any paid employees, in both retail and manufacturing. When we adjust

the reported revenue in the household business sector by this factor, the productivity gap falls from 6.0 to 4.2 (column 1, row 2 of Table B.5).

Second, people working in the household business sector might overstate effective hours worked. For example, a shopkeeper might be watching her children while tending to the business, but reports total hours worked in the shop in the survey. We are not aware of a study that examines potential mismeasurement of labor supply to microenterprises. However, the data from Ghana generated by Fafchamps, McKenzie, Quinn and Woodruff (2014), which covers a sample of microenterprises without paid employees, asked about the number of hours worked last week as well as the number of hours worked at full effort.[9] Interestingly, the difference in hours worked versus full effort hours is not very large. Across all rounds of the surveys, microenterprise owners reported working at full effort 89 percent of the time. This suggests that reports of hours worked may slightly overestimate true labor input. When we adjust the productivity gap for the measurement error in revenue and hours worked, the ARPL gap is 3.7 (column 1, row 3 of Table B.5).

B.1.5. Measurement issues specific to the 2002 VHLSS

The above discussion focuses on measurement issues related to informal businesses in any data. In this section, we use the 2004 and 2006 VHLSSs data to assess two measurement issues specific to the 2002 VHLSS data used in the current study.

First, our estimate of aggregate revenue in the household business sector is likely overestimated due to the inclusion of some private enterprises in the estimate. The 2002 VHLSS asked each household whether they ran a business,

[9] This question was asked because many microenterprise owners are simultaneously engaging in non-microenterprise activities, such as childcare or household work, while operating their business. We thank Christopher Woodruff for making us aware of this data.

but it did not distinguish between household businesses and private enterprises. Consequently, revenue from private enterprises is also included in our estimate of aggregate revenue in the household business sector. However, the labor module clearly distinguishes between working in a household business versus a private enterprise. Hence, our estimate of ARPL in the household business sector is an overestimate, as it is based on the revenue from household businesses and private enterprises. This underestimates the ARPL gap between the enterprise and household business sectors.

The 2006 VHLSS distinguished between the two types of private businesses, household businesses and private enterprises, and we use this survey to get an estimate of how much we might be overstating aggregate revenue from the household business sector by including private enterprises. The estimate based on the 2006 data likely overstates the contribution of private enterprises to the 2002 household business revenue because the relative share of the formal sector has been growing over time. As a result, the adjustments below should be viewed as providing a lower-bound estimate of the ARPL. Only 1.5 percent of manufacturing businesses in the 2006 VHLSS were private enterprises. However, they tend to be bigger. In the 2006 VHLSS, among manufacturing businesses, 23 percent of aggregate revenue is from private enterprises. Under the assumption that this share has not changed over time, we adjust the ARPL ratio for this measurement error by subtracting the same proportion of revenue from the 2002 VHLSS based estimates. The ARPL gap rises to 4.9 (column 1, row 4 of Table B.5).

In addition, our estimate of hours worked in the household business sector omits hours worked as a secondary job because the 2002 VHLSS did not collect detailed data on secondary jobs, defined as the second most time-consuming job during the past 12 months. This underestimates labor supplied to the household business sector. We assess the potential measurement error with the 2004 and

2006 VHLSSs, which asked detailed questions about both the primary and secondary jobs of workers. In the enterprise sector, only 1.2 percent and 1.0 percent of total hours reported in the 2004 and 2006 VHLSSs were worked in secondary jobs. However, 13.9 percent and 13.8 percent of total hours in the household business sector in the 2004 and 2006 VHLSSs were worked as a secondary job. Adjusting the 2002 estimate of the ARPL ratio to include the missing hours in the household business sector due to secondary jobs increases the estimate to 5.6 (column 1, row 5 of Table B.5).

In summary, while measurement error is certainly present in our data, attempts to adjust for plausible measurement error do not eliminate the ARPL gap across the household business and enterprise sectors within manufacturing. In the main text we focus on the most conservative estimate of the productivity gap from column 1 of Table B.5, namely 3.7.

B.1.6. Differences in the output elasticity of labor

As equation (4) in Section V makes clear, the ARPL gap may overestimate the MRPL gap if the household business sector has a higher output elasticity of labor, because $\frac{MRPL_e}{MRPL_h} = \frac{\left(1-\alpha_e\right)}{\left(1-\alpha_h\right)}\frac{ARPL_e}{ARPL_h}$. Estimating production functions is beyond the scope of this paper. Hence, we use existing values from the literature. There are not many studies from which to draw. Fernandez and Meza (2015) calibrate a model of informal and formal firms using output elasticities of labor of 0.8 and 0.65, respectively. Restrepo-Echavarria (2014) assumes output-labor elasticities of 1 and 0.68 in the informal and formal sectors, respectively. [10] We choose the largest elasticity ratio of about 1.5 (approximately

[10] Restrepo-Echavarria (2014) reports value added-labor elasticity. Adjusting for factor share of materials based on Nataraj (2011) yields similar results because the factor share of materials is very similar across the two sectors.

1/0.68) and apply it to the ARPL gaps reported in column 1 of Table B.5. The obtained MRPL ratios are reported in column 3 of Table B.5.

Let us first focus on the ARPL gap that adjusts for worker heterogeneity and hours worked differences across the two sectors, reported in row 1. Adjusting this ratio with the elasticity ratio suggests an MRPL ratio of about 4 (row 1, column 3 of Table B.5). MRPL ratios associated with alternative estimates of labor productivity gaps in column 1 of Table B.5, discussed in Section B.1.5, are reported in the remaining rows in column 3. The most conservative MRPL ratio of 2.5 is the one reported in row 3 and is based on the productivity gap of 3.7 that adjusts for worker heterogeneity, and the measurement error in revenue and hours worked. Note that we used a factor of 1.5, a very large difference in elasticities, to provide a conservative estimate of the possible MRPL gap. In the main text, we therefore discuss this most conservative estimate as the lower bound for the labor productivity gap, adjusted for measurement error and differences in output-labor elasticities across sectors.

B.2. Estimating the share of workers reallocated

The estimate of the aggregate labor productivity gain in Section V of the main text requires an estimate of the share of manufacturing workers reallocated from the household business to the enterprise sector due to the BTA, s^{BTA}.

Within each industry j, we estimate the share of workers reallocated as $\hat{\beta} \times \Delta tariff_j$, where $\hat{\beta}$ is estimated based on equation (1) and $\Delta tariff_j$ is the change in the U.S. tariff on Vietnamese exports in industry j due to the BTA. We then sum over manufacturing industries, weighting by the industry's share of overall manufacturing employment:

$$\sum_j \left(\hat{\beta} \times \Delta tariff_j \right) s_j$$

where s_j is the share of manufacturing workers in industry j. To be consistent with how we calculate the number of workers for the ARPL gaps, we calculate total employment in each industry as the sum of household business sector employment, estimated from the 2002 VHLSS, and enterprise sector employment, derived from employment at the end of 2001 from the enterprise data.[11]

Our preferred estimate for $\hat{\beta}$ is 0.156 (see Panel A, column 3 of Table 3). With this value, we estimate that 4.9 percent of manufacturing workers were reallocated out of the household business sector to the enterprise sector. The regression estimates reported in Table 3 focus on the reallocation of workers between sectors, whereas Section V focuses on labor productivity per hour worked, not per worker. In our case, this distinction in the reallocation of labor, whether workers or hours, turns out to be inconsequential as the BTA induced essentially an identical change in the share of hours worked in the household business sector (see the coefficient of 0.157 in column 3 of Table A.5 for share of hours worked as compared to our benchmark estimate of 0.156 from Table 3). The estimates of reallocated hours worked are reported in Table 8 in the main text. They are used in all subsequent calculations discussed in Section B.3.

B.3. Aggregate Labor Productivity Gain

We use the formula in Section V of the main text, equation (5), to compute the gain in labor productivity based on various labor productivity gaps. Recall that these labor productivity gaps are reported in Appendix Table B.4 and column 1 and 3 of Table B.5. Those tables also report the associated annual aggregate labor productivity gaps.

Here we focus our discussion on the estimates of the aggregate gains for the labor productivity gaps reported in Appendix Table B.5. Recall that in column

[11] If we estimate industry employment solely from the 2002 VHLSS we arrive at a similar estimate of the share of manufacturing workers reallocated.

1, row 1, we present an ARPL gap of 6.0, which adjusts for worker heterogeneity. Subsequent rows sequentially adjust the ARPL gap for each measurement error issue discussed previously. The ARPL gap falls as we consider possible reporting error in household business revenue and labor input and then increases as we remove possible revenue from private enterprises being included as household business revenue and add additional labor inputs from secondary jobs. These movements in the ARPL gap are reflected in the estimates of aggregate labor productivity gains. Our preferred labor productivity gap estimate of 3.7 in row 3 suggests that aggregate labor productivity per hour worked increased by 2.8 percent annually over the period of the study. This is the estimate also reported in Table 8 and discussed in Section V of the main text. We view the 3.5 percent estimate in row 1, column 2 (which is based on the labor productivity gap that only adjusts for worker heterogeneity) as the upper bound on the aggregate labor productivity gains. Overall, the estimated productivity gains range from 2.8 to 3.5 percent per year.[12]

Column 4 of Appendix Table B.5 presents estimates of the annual growth in labor productivity that take into account potential measurement issues and additionally allow for differences in the output-labor elasticity across the household business and enterprise sectors. These are based on the MRPL gaps reported in column 3. The growth estimates range from 1.5 to 2.1 percent annually. Our most conservative estimate of the aggregate productivity gain is 1.5 percent (row 3, column 4). We discuss this gap in the main text and Table 8.

In Appendix Table B.6, we replicate the analysis of Table 8, except that labor productivity is based on per worker instead of per hour worked. We do so because micro data on hours worked is often not available in firm-level datasets

[12] In Appendix Table B.4, we also report growth in aggregate labor productivity based on the labor productivity gap using data from four years later. Although the labor productivity gap is lower, 3.5 versus 6.0, after adjusting for differences in human capital and hours worked, the associated growth in aggregate labor productivity is 2.7 percent per year.

and thus labor inputs are typically measured on a per worker basis. We focus our discussion on column 1, which is based on all of manufacturing. The unadjusted labor productivity gap is 9, as in Table 8. Subsequently, we adjust for differences in human capital, but not in hours worked, and estimate a labor productivity gap of 7.5. Additional adjustments for measurement error in revenue and labor inputs for household businesses as well as the difference in the output-labor elasticity across sectors are the same as in Table 8 for labor productivity per hour. At the bottom of column 1 in Table B.6, we report the aggregate labor productivity estimates due to the BTA associated with the reported labor productivity gaps. These range between 2.0 and 4.1 percent annually, with an estimate of 3.4 percent as our preferred estimate of the gains in aggregate labor productivity per worker. This preferred estimate parallels that for aggregate labor productivity per hour in that the associated labor productivity gap in both cases is based on adjustments for measurement error in revenue and labor inputs for household businesses, but assumes the same output-labor elasticity across sectors.

We can also use equation (5) to estimate the growth in wages due to the BTA-induced reallocation of workers towards formal firms within manufacturing. To do so, we simply use the labor productivity gap based on the wage gap and use initial wage levels in each sector instead of labor productivity levels. The wage gap adjusted for worker heterogeneity and annual hours worked is reported in row 2, column 2 of Table. The associated growth in wages per hour worked is 0.5 percent annually. In Appendix Table B.6, we report the wage gap based on annual earnings (i.e., not adjusting for differences in hours worked between the informal and formal sectors), but adjusted for differences in human capital (row 2, column 2). The associated gain in annual wages is 0.9 percent per year.

Appendix Table B.1: Differences in hourly wages between enterprise and household business sector workers in manufacturing

Dependent variable: ln real hourly wage

	(1)	(2)	(3)	(4)	(5)	(6)
Enterprise sector indicator	0.366***	0.296***	0.291***	0.160***	0.227***	0.221***
	(0.0129)	(0.0133)	(0.0133)	(0.0131)	(0.0127)	(0.0136)
Observations	9,416	9,416	9,416	9,416	9,416	9,416
R-squared	0.079	0.101	0.104	0.280	0.356	0.372
Location?	No	No	No	Yes	Yes	Yes
Demographics?	No	No	No	No	Yes	Yes
Years of education?	No	Yes	No	No	No	No
Education categories?	No	No	Yes	Yes	Yes	Yes
Industry FEs?	No	No	No	No	No	Yes

Notes: ***, **, and * denotes significance at 1, 5, and 10 percent level, respectively. The sample is all wage workers in manufacturing age 20 to 64 in the 2002 VHLSS. Location controls include a rural identifier and province fixed effects. Demographic controls include age, age squared, a female indicator, and an ethnic minority indicator. Education controls include indicators for completed primary, completed lower secondary, and completed upper secondary with did not complete primary as the excluded category. Industry controls include industry fixed effects.

Appendix Table B.2: Differences in hourly wages between enterprise and household business sector workers in manufacturing

Dependent variable: ln real hourly wage

	(1)	(2)	(3)	(4)
Enterprise sector indicator	0.0874*	0.0892*	0.202***	0.191***
	(0.0459)	(0.0463)	(0.0234)	(0.0354)
Observations	2,680	2,680	2,680	1,340
R-squared	0.052	0.077	0.435	0.445
Number of individuals	1,340	1,340	1,340	1,340
Location?	No	No	Yes	Yes
Demographics?	No	No	Yes	Yes
Years of education?	No	No	No	No
Education categories?	No	No	Yes	Yes
Industry?	No	Yes	Yes	Yes
Year FEs?	Yes	Yes	Yes	No
Individual FEs?	Yes	Yes	No	No

Notes: ***, **, and * denotes significance at 1, 5, and 10 percent level, respectively. The sample is a panel of workers age 20 to 64 in the 2002 VHLSS that worked in manufacturing in 2002 and 2004 for a wage. In column (4) the analysis is restricted to 2002.

Appendix Table B.3: Differences in hourly earnings between enterprise and household business
Dependent variable: ln real hourly earnings

	(1)	(2)	(3)	(4)	(5)	(6)
Enterprise sector indicator	0.304***	0.235***	0.225***	0.0713***	0.126***	0.112***
	(0.0126)	(0.0131)	(0.0131)	(0.0130)	(0.0123)	(0.0133)
Observations	12,943	12,943	12,943	12,943	12,943	12,943
R-squared	0.039	0.059	0.062	0.219	0.288	0.302
Location?	No	No	No	Yes	Yes	Yes
Demographics?	No	No	No	No	Yes	Yes
Years of education?	No	Yes	No	No	No	No
Education categories?	No	No	Yes	Yes	Yes	Yes
Industry FEs?	No	No	No	No	No	Yes

Notes: ***, **, and * denotes significance at 1, 5, and 10 percent level, respectively. The sample is all wage workers and managers of one-worker household businesses in manufacturing age 20 to 64 in the 2002 VHLSS. Location controls include a rural identifier and province fixed effects. Demographic controls include age, age squared, a female indicator, and an ethnic minority indicator. Education controls include indicators for completed primary, completed lower secondary, and completed upper secondary with did not complete primary as the excluded category. Industry controls include industry fixed effects.

Appendix Table B.4: Productivity gap between the enterprise and household business sectors in manufacturing in 2005

	Average revenue product of labor			
	Manuf. (2006 VHLSS) (1)	Manuf. (2)	Textiles and apparel (3)	Ho Chi Minh City and Dong Nai (4)
Productivity gap	3.6	5.0	5.0	3.3
Productivity gap adjusted by hours worked & human capital	2.8	3.5	3.9	2.6
Share of hours reallocated to enterprises due to the BTA	0.050	0.050	0.086	0.053
Initial share of hours in the household business sector	0.597	0.597	0.615	0.380
Annual growth in revenue per hour worked (%)	2.4	2.7	5.3	1.8

Notes: The productivity gap for the average revenue product of labor is the ratio of revenue per worker in the enterprise sector to revenue per worker in the household business sector. In column 1, we report the ratio for private enterprises versus household businesses as calculated from the business module in the 2006 VHLSS. In columns 2 through 4, we report the ratio for all enterprises, as calculated using the 2005 enterprise data, versus household businesses from the 2006 VHLSS. See section V and Appendix B for further details on the calculations and data sources.

Appendix Table B.5: Adjusting the average revenue product of labor gap

Adjusted for:	Measurement error		Measurement error and output-labor elasticity	
	ARPL gap	Annual growth rate (%)	MRPL gap	Annual growth rate (%)
Adjusted for hours and human capital	6.0	3.5	4.0	2.1
+ Adjusted for underreporting of microenterprise revenue by 30%	4.2	3.0	2.8	1.7
+ Adjusted for hours worked at full effort	3.7	2.8	2.5	1.5
+ Adjusted for revenue from included private enterprises	4.9	3.3	3.2	2.0
+ Adjusted for hours worked in a secondary job	5.6	3.5	3.8	2.1

Notes: The table reports the ARPL gap between the enterprise and household business sectors adjusted for measurement error and differences in output-labor elasticities across sectors. Each row is based on the adjustment described in the row and all adjustments in previous rows. It also reports the annual growth rate of hourly labor productivity associated with each adjusted ARPL gap.

Appendix Table B.6: Productivity gap per worker between the enterprise and household business sectors in manufacturing

	Manufacturing		Textiles and apparel		Ho Chi Minh City and Dong Nai	
	Revenue based (1)	Wage based (2)	Revenue based (3)	Wage based (4)	Revenue based (5)	Wage based (6)
Productivity gap						
Unadjusted	9.0	1.8	6.6	1.7	7.0	1.5
Adjusted by human capital	7.5	1.5	5.8	1.5	5.9	1.2
+ measurement error in revenue and hours worked	4.6		3.6		3.7	
+ differences in output-labor elasticity	3.1		2.4		2.4	
Share of workers reallocated to enterprises due to the BTA	0.049	0.049	0.085	0.085	0.053	0.053
Initial share of workers in the household business sector	0.648	0.648	0.664	0.664	0.395	0.395
Annual growth (%)						
Adjusted by human capital	4.1	0.9	6.9	1.7	2.7	0.4
+ measurement error in revenue and hours worked	3.4		5.4		2.3	
+ differences in output-labor elasticity	2.0		2.9		1.2	

Notes: The labor productivity gap reported in columns 1, 3, and 5 is based on the average revenue product of labor and subsequent adjustments. The average revenue product of labor is the ratio of revenue per worker in the enterprise sector to revenue per worker in the household business sector. The labor productivity gap reported in columns 2, 4, and 6 is based on the ratio of annual earnings per worker in the enterprise sector to annual earnings per worker in the household business sector, plus subsequent adjustments. The difference in output-labor elasticity allow the MRPL and ARPL gaps to differ. See section V and Appendix B for further details on the calculations and data sources.