An econometric analysis of arms imports

RON SMITH[a;] , ALI TASIRAN[a;b]
[a]Birkbeck College, University of London
[b]Göteborg University

June 2003

Abstract

This paper examines the determinants of arms imports for a panel of 52 countries 1981-1999 where there are non-zero observations for both the main measures, SIPRI and WMEAT. It also discusses the characteristics of the two series and estimation and imputation methods for a larger sample where there are zeros and missing observations. This is very preliminary and incomplete work in progress, please do not cite without the author's permission.We would be grateful for comments to R.Smith@bbk.ac.uk.

J.E.L. Classi...cation: C23, C24, D74

Acknowledgement 1 We are very grateful to various members of SIPRI for help with the data. Any opinions expressed here are our own and are not necessarily those of SIPRI or the Bureau of Veri...cation and Compliance of the US Department of State whose data we have used.

1 Introduction

The arms trade is interesting because it is where foreign policy concerns such as security, human rights and international order interact with economic concerns such as trade, jobs and pro...ts. Levine and Smith (1997, 2000) provide a review of the issues. There is also concern that the arms trade incites political violence e.g. Craft and Smaldone (2002). However, there has been relatively little work on the determinants of arms imports, as compared to a

large amount of work on the determinants of military expenditures. There are two sources of data on arms imports: SIPRI and WMEAT (previously published by ACDA now by BVC). SIPRI measures the volume of transfers of major weapons systems: quantities are multiplied by trend indicator values of unit production costs, irrespective of the price actually paid. Thus when Germany transferred most of the old East German navy to Indonesia virtually free, this is still a quantity of weapons transferred and will be re‡ected as such by the SIPRI measure. WMEAT measures the value of arms transfers, re‡ecting the price actually paid. There are thus two di¤erences between the series: valuation (SIPRI is a quantity measure WMEAT a value measure) and coverage (SIPRI covers major weapons systems, WMEAT all systems). The approach in this paper extends the analysis in Levine, Mouzakis and Smith (1998), henceforth LMS. LMS ignore the coverage di¤erence and note that the ratio of a value series to a volume series produces a price index, thus one could make the quantity of arms imported a function of price and other variables and measure the price elasticity of demand for arms imports. LMS used a cross-section of 38 countries in 1990-1991. This paper uses a much larger panel data set for the period 1981-1999, allowing us to investigate both the time-series and cross-section variation and relax a number of assumptions in LMS. Explaining arms imports is more di¢cult than explaining military expenditure because the data are very lumpy, imports can be small in one year, large in the next as say a squadron of aircraft are delivered and go back to being small. This suggests a process of non-linear stock adjustment. There are also a lot of zeros in the data, where countries did not import at all. In addition, there are missing observations and measurement errors and cases where SIPRI record an import and WMEAT does not and vice versa. Thus the data raise a number of interesting technical statistical questions. There are data of a sort for 174 countries listed in the appendix. Some of these ceased to exist, e.g. the Soviet Union, others are recent creations, e.g. Russia. Some are very small and in many cases data on arms imports are di¢cult to obtain. Thus we look at sub-sets of di¤erent quality, recognising the trade-o¤ between ignoring information and contaminating our estimates with very inaccurate information.

Section 2 provides a brief review of the panel estimators that we shall use to establish terminology. We begin by considering a panel of panel of 52 countries for which there are non-zero observations on both the SIPRI and WMEAT measure for the years 1981-1999. This avoids the zero and missing observations problems. Section 3 discusses the data on arms imports and the

measurement issues for this sample. Section 4 provides some initial estimates of demand comparable to LMS using this sample. The remaining sections describe our research approach, rather than present results. This is very much work in progress. The next stages are: (1) to examine non-linear stock adjustment for this sample; (2) extend the sample to analyse the pattern of zeros and missing observations (3) examine imputation measures that can be used when we have observations from one source but not the other (4) simultaneously model the zero/non-zero decision (whether to import or not) and the non-linear stock adjustment decision (how much to import given that a country does import). In each case we plan to examine the robustness of the results to di¤erent estimators.

2 Panel Estimators

Suppose that we have data on a dependent variable y_{it}; (arms imports in this case) for a sample of countries $i = 1; 2; :::; N$ and years $t = 1; 2; :::; T$ and data on a vector of k independent variables x_{it}: De…ne the country and overall means as

$$\bar{y}_i = T^{i\,1} \sum_{t=1}^{T} y_{it};\ \bar{y} = (NT)^{i\,1} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}$$

then the total variation in y_{it} is the sum of the within country variation and the between country variation

$$\sum_i \sum_t (y_{it} \text{ i } \bar{y})^2 = \sum_i \sum_t (y_{it} \text{ i } \bar{y}_i)^2 + T \sum_i (\bar{y}_i \text{ i } \bar{y})^2$$

similarly for x_{it}: The main panel estimators we shall use are

(1) pooled OLS which gives the within and between variation equal weight

$$y_{it} = \alpha + \beta' x_{it} + u_{it};$$

(2) the within estimator (also known as the one way …xed e¤ects, least squares dummy variables and a variety of other names) which allows intercepts to di¤er and constrains the slopes to be the same

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it}$$

this only uses the within variation and is equivalent to OLS on

$$(y_{it} - \bar{y}_i) = \gamma'(x_{it} - \bar{x}_i) + u_{it};$$

(3) the two way fixed effects estimator which constrains slopes to be the same, but allows intercepts to vary freely both over country and year

$$y_{it} = \alpha_i + \alpha_t + \gamma' x_{it} + u_{it}$$

this allows for a completely flexible trend which impacts each country by the same amount;

(4) the between estimator, which only uses the cross-section variation in the country means

$$\bar{y}_i = \alpha + \gamma'\bar{x}_i + \bar{u}_i;$$

(5) The random coefficient model which allows all the parameters to differ, i.e.

$$y_{it} = \alpha_i + \gamma_i' x_{it} + u_{it}$$

and calculates weighted averages of the individual estimates $\hat{\beta}_i$; the weights being based on the variances of the $\hat{\beta}_i$:

$$\tilde{\beta} = \sum_i W_i \hat{\beta}_i.$$

There is another estimator, which we shall not use, the random effects estimator, which assumes that slopes are identical and intercepts are randomly distributed independently of the regressors. It calculates the optimal combination of within and between variation under these assumptions. In static models like these, if the coefficients, α_i and γ_i are randomly distributed, independently of the x_{it} all these estimators will produce unbiased estimators of the expected values of the coefficients $E(\alpha_i)$ and $E(\gamma_i)$: However, the independence assumption may not hold and these estimators can be very different, in particular the cross-section (between country) effects can be very different from the time-series (within country effects). In figure 1, within each country the relationship is negative, but the intercepts are positively correlated with the mean level of x, thus the between cross-section relationship is positive. In figure 2, the slope γ_i is negatively correlated with the mean of x giving the appearance of a quadratic relationship. Comparison of the estimators allows us to compare the within and between relationships and judge how robust the estimates are. (FIGURES TO BE ADDED).

A further issue arises with dynamic models since the within (…xed e¤ect) estimator of

$$y_{it} = \alpha_i + \beta' x_{it} + \gamma y_{it-1} + u_{it}$$

is consistent for $T \to \infty$; but is not consistent for $N \to \infty$; T small. In this case $\hat{\gamma}$ is biased downwards. This is the standard small T bias of the OLS estimator of models with lagged dependent variables. There are a variety of instrumental variable estimators for this case. However, if the true model is heterogeneous

$$y_{it} = \alpha + \beta_i' x_{it} + \gamma_i y_{it-1} + u_{it}$$

and homogeneity of the slopes is incorrectly imposed, the within estimator is not consistent $T \to \infty$ and $\hat{\gamma}$ is biased upwards towards unity (assuming x_{it} is positively serially correlated as is usually the case). The RCM estimator is however consistent $T \to \infty$; though it also su¤ers the small T bias from the lagged dependent variable. Comparison of the various estimators, which are subject to di¤erent biases, e.g. through Hausman tests, can allow us to infer which biases are most important.

3 The measures of arms imports

There are two sources of data on arms imports: SIPRI and WMEAT (previously published by ACDA now by BVC). SIPRI measures the volume of transfers of major weapons systems: quantities are multiplied by trend indicator values of unit production costs, irrespective of the price actually paid. WMEAT measures the value of arms transfers, re‡ecting the price actually paid. There are thus two di¤erences between the series: valuation (SIPRI is a quantity measure WMEAT a value measure) and coverage (SIPRI covers major weapons systems, WMEAT all systems). If there were no measurement errors the we could say the SIPRI measure of imports by country i in year t; $S_{it} = Q^M_{it}$ while the WMEAT measure $A_{it} = P_{it}(Q^M_{it} R_{it})$; where $R_{it} = Q_{it}/Q^M_{it}$ total quantity divided by the quantity of major weapons systems. The ratio

$$\frac{A_{it}}{S_{it}} = P_{it} R_{it}$$

LMS suggested ignoring the term R_{it} and using the ratio of the WMEAT series to the SIPRI series as a measure of price. With a panel of data, we can investigate the issues a little bit more. We took the 52 countries for

which there were non-zero observations of arms imports by both SIPRI and WMEAT in every year of the period 1981-1999 and also full information in WMEAT on military expenditure, number in the armed forces, real GDP and population. The Countries are Algeria, Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, China, Taiwan, Colombia, Denmark, Ecuador, Egypt, Finland, France, Germany, Greece, India, Indonesia, Iran, Israel, Italy, Japan, Jordan, Kenya, S. Korea, Kuwait, Malaysia, Mexico, Morocco, Netherlands, Norway, Oman, Pakistan, Peru, Phillipines, Saudi Arabia, Singapore, South Africa, Spain, Sudan, Sweden, Switzerland, Syria, Thailand, Tunisia, Turkey, UAE, UK, USA, Venezuela. We return to the issue of zero observations below.

Taking logs of the relationship and using lower case letters to denote logarithms we have

$$a_{it} - s_{it} = p_{it} + r_{it} = u_{it}$$

Above we have assumed that they are measured without error, but any measurement errors could be regarded as absorbed into the term d_{it}: Assume that the variables are measured as deviations from their means. Since the variables are logarithms this will remove any scale effect from units of measurement. Consider the two possible regressions of the SIPRI measure on the WMEAT and the reverse

$$\begin{aligned} s_{it} &= b_1 a_{it} + e_{1;it} \\ a_{it} &= b_2 s_{it} + e_{2;it} \end{aligned}$$

The underlying data generation process is such that $b_1 = b_2 = 1$; $e_{1;it} = -u_{it}$; $e_{2;it} = u_{it}$: This will not be revealed by the regressions since both a_{it} and s_{it} are likely to be correlated with u_{it}; if only because quantity s_{it} will be a function of price. Suppose we have

$$\begin{aligned} s_{it} &= -\bar{p}_{it} + \varepsilon_{it} \\ a_{it} &= s_{it} + p_{it} + r_{it} \end{aligned}$$

with ε_{it} and p_{it} is independent of ε_{it} and r_{it} is independent of s_{it}; p_{it}: Denote

the variances as $\sigma_{\epsilon\epsilon}$; σ_{rr}; σ_{pp}; then

$$\begin{aligned}
\sigma_{ss} &= \gamma^2\sigma_{pp} + \sigma_{\epsilon\epsilon} \\
\sigma_{sp} &= -\gamma\sigma_{pp} \\
\sigma_{aa} &= \sigma_{ss} + \sigma_{pp} + \sigma_{rr} + 2\sigma_{sp} \\
&= \sigma_{rr} + \sigma_{\epsilon\epsilon} + (\gamma - 1)^2\sigma_{pp} \\
\sigma_{as} &= \sigma_{ss} + \sigma_{sp} \\
&= \sigma_{\epsilon\epsilon} + \gamma(\gamma - 1)\sigma_{pp}
\end{aligned}$$

We observe σ_{as} σ_{ss} σ_{aa}. The sums of squares are given below.

Sums of Squares	LS	LA	LA − LS
Total	2420	1796	1192
Within	1156	588	922

The variance of $a - s = \sigma_{aa} + \sigma_{ss} - 2\sigma_{sa}$:

Our regression estimates are

$$\begin{aligned}
b_1 &= \frac{\sigma_{sa}}{\sigma_{aa}} = \frac{\sigma_{ss} - \gamma\sigma_{pp}}{\sigma_{ss} + \sigma_{pp} + \sigma_{rr} - 2\gamma\sigma_{pp}} = \frac{\sigma_{ss} - \gamma\sigma_{pp}}{\sigma_{ss} + \sigma_{rr} + \sigma_{pp}(1 - 2\gamma)} \\
&= \frac{\sigma_{\epsilon\epsilon} + \gamma(\gamma - 1)\sigma_{pp}}{\sigma_{rr} + \sigma_{\epsilon\epsilon} + (\gamma - 1)^2\sigma_{pp}} \\
b_2 &= \frac{\sigma_{sa}}{\sigma_{ss}} = \frac{\sigma_{ss} - \gamma\sigma_{pp}}{\sigma_{ss}} = \frac{\sigma_{\epsilon\epsilon} + \gamma(\gamma - 1)\sigma_{pp}}{\sigma_{\epsilon\epsilon} + \gamma^2\sigma_{pp}}
\end{aligned}$$

Notice that $b_1 b_2 = R^2$; the coefficient of determination for either of the regressions. We observe two variances and a covariance, which depend on four unknown parameters so the model is not identified. However, it can give us a feel for the sort of effects operating.

If $\gamma = 0$; we have

$$b_1 = \frac{\sigma_{\epsilon\epsilon}}{\sigma_{rr} + \sigma_{\epsilon\epsilon} + \sigma_{pp}}; \quad b_2 = 1:$$

The result $b_2 = 1$; arises because the error $(p + r)$ is not correlated with a:

If $\gamma = 1$; we have

$$b_1 = \frac{\sigma_{\epsilon\epsilon}}{\sigma_{rr} + \sigma_{\epsilon\epsilon}}; \quad b_2 = \frac{\sigma_{\epsilon\epsilon}}{\sigma_{\epsilon\epsilon} + \sigma_{pp}}$$

so $b_1 > b_2$ if $\sigma_{pp} > \sigma_{rr}$; the price e¤ect in the SIPRI measure is larger than the error. By adding other variables, as we do in the next section, we can reduce σ_{rr}: We might also model the di¤erence between the measures, e.g. assume that $p_{it} + r_{it} = \alpha_i + \alpha_t$ if this was the case then our estimates of b_1 in the regression of SIPRI on WMEAT would be closer to unity in a two way …xed e¤ect model.

The regressions were estimated four ways, simple pooled OLS, the one way …xed e¤ect estimator, the two way …xed e¤ect estimator and the between estimator. First the regression of s_{it} on a_{it}

Estimator	b_1	R^2
	(se)	SER
OLS	0:84	0:52
	(0:02)	1:078
Within	0:699	0:64
	(0:04)	0:96
two ¡ way	0:75	0:67
	(0:04)	0:93
between	0:91	0:79
	(0:06)	0:52

then the reverse regression of a_{it} on s_{it}

Estimator	b_2	R^2
	(se)	SER
OLS	0:62	0:52
	(0:02)	0:93
Within	0:35	0:75
	(0:02)	0:69
two ¡ way	0:38	0:78
	(0:02)	0:66
between	0:87	0:79
	(0:06)	0:51

The R^2 for the OLS and the between have to be the same for the two regressions, this in not the case for the other two estimators. In both cases the between estimator is closer to unity and not signi…cantly di¤erent from unity in the case of b_1: The within estimates are the lowest, the pooled lie between the within and between (they average both), the two way estimates

are higher than the within, but lower than the OLS or between. Whatever the estimator the estimate for b_1 is closer to unity than b_2:

4 Demand Functions

We will proceed on the basis that it is reasonable to treat the ratio of the WMEAT to the SIPRI series as a noisy measure of price and estimate demand functions. The demand function reported in LMS was

$$s_i = \bar{\ }_0 \text{ ¡ } 0{:}56p_i + 1{:}52m_i \text{ ¡ } 0{:}06m_i^2 + 4{:}74ypc_i \text{ ¡ } 0{:}29ypc_i^2 + u_i$$

though only military expenditure (t=2.2) and price (t=-3.2) were signi…cant, (the standard error for price given in the paper 0.71 should be 0.17). The income measure was the Penn World Table PPP per-capita income. The argument for a non-monotonic relationship was that countries with low levels of military expenditure do not import because they do not need the weapons; countries with very high levels of military expenditure do not import because they have their own arms industry; so we expect imports to rise and then fall with military expenditure, which these estimates show. We replicated this on our panel using the same four estimators as above but the WMEAT measure of per-capita income, coe¢cients signi…cant at the 5% level are shown in bold.

	p	m	m^2	ypc	ypc^2	R^2 SER
OLS	¡ 0:86	2:41	¡ 0:01	¡ 1:21	0:03	0:57
	(0:03)	(0:19)	(0:01)	(0:58)	(0:02)	1:02
within	¡ 0:81	¡ 0:38	0:06	¡ 1:32	0:04	0:79
	(0:02)	(0:59)	(0:04)	(0:88)	(0:03)	0:74
two ¡ way	¡ 0:81	¡ 0:46	0:07	¡ 0:76	0:02	0:79
	(0:03)	(0:60)	(0:04)	(0:93)	(0:03)	0:74
between	¡ 1:05	2:39	¡ 0:10	¡ 1:19	0:03	0:57
	(0:25)	(0:71)	(0:04)	(2:31)	(0:07)	0:75

The year e¤ects were not signi…cant. Price is always signi…cant, with quite similar estimates, the between estimate somewhat larger. Military expenditure is positive and signi…cant in the OLS and between, but is negative, though not signi…cantly so for the within and two way, military expenditure

squared is the opposite sign to military expenditure but only signi…cant in the between. Income has a negative coe¢cient, which is quite similar between the estimators, but is only signi…cant in OLS, income squared always has a positive coe¢cient but is never signi…cant. The most interesting feature is that military expenditure is having a quite di¤erent e¤ect in cross-section (where the inverted U shaped pattern identi…ed in LMS is still apparent) and time-series.

The models were re-estimated using total GDP rather than per-capita.

	p	m	m^2	y	y^2	R^2 SER
OLS	¡0:86	2:35	¡0:08	¡0:27	0:005	0:59
	(0:03)	(0:29)	(0:02)	(0:38)	(0:02)	1:00
within	¡0:80	1:62	¡0:06	¡4:68	0:20	0:80
	(0:02)	(0:64)	(0:04)	(0:63)	(0:03)	0:73
two ¡ way	¡0:80	1:62	¡0:06	¡4:64	0:20	0:80
	(0:03)	(0:65)	(0:04)	(0:64)	(0:03)	0:73
between	¡1:21	1:59	¡0:03	0:76	¡0:06	0:68
	(0:22)	(0:98)	(0:06)	(1:27)	(0:05)	0:68

Again the time e¤ects were not signi…cant. Use of total GDP has produced a slight improvement for all except the between regression, where nothing except price is now signi…cant. The sign pattern of the military expenditure estimate is now the same for each estimator.

Another advantage of panel data is that one can allow for dynamics. After some experimentation the following within estimates were obtained

$$s_{it} = \underset{(0:02)}{¡0:82p_{it}} \; \underset{(0:03)}{+0:50p_{it¡1}} \; \underset{(0:07)}{+0:35m_{it}} \; \underset{(0:03)}{+0:57s_{it¡1}}$$

This had an R^2 = 0:86 and SER = 0:60: Time e¤ects were not signi…cant, nor was income or lagged or squared military expenditure. The long-run e¤ect of prices is (¡0:82 + 0:50)=(1 ¡ 0:57) = ¡0:74; the long-run e¤ect of military expenditure is 0:35=(1 ¡ 0:57) = 0:81: The coe¢cient of the lagged dependent variable in the within estimator is biased upwards if the slope coe¢cients are not in fact equal. The extent of this bias can be checked using a random coe¢cient estimator. The estimates for this, which average

the individual equations were:

$$s_{it} = \underset{(0:05)}{¡0:82p_{it}} \quad \underset{(0:05)}{+0:44p_{it¡1}} \quad \underset{(0:24)}{+0:53m_{it}} \quad \underset{(0:05)}{+0:51s_{it¡1}}$$

The long-run price e¤ect of ¡0:76 is quite similar, the long-run military expenditure e¤ect of 1:08 is rather larger than given by the within estimator: There is an issue of endogeneity, so the equation was estimated by instrumental variables (two stage least squares). As instruments were used the predetermined variables in the equation, $p_{it¡1}$; m_t; $s_{it¡1}$; y_{it} and y^2_{it}; re‡ecting the possibility that the price charged would depend on the income of the country and time dummies to capture shifts in the price over time, these were signi…cant. The IV within estimates were:

$$s_{it} = \underset{(0:40)}{¡1:28p_{it}} \quad \underset{(0:15)}{+0:67p_{it¡1}} \quad \underset{(0:09)}{+0:31m_{it}} \quad \underset{(0:03)}{+0:58s_{it¡1}}$$

with an $R^2 = 0:79$ and $SER = 0:73$: The long-run e¤ect of price is ¡1:45 and of military expenditure 0:73: If price were signi…cant just because it was constructed from the dependent variable s_{it}; one would expect the IV estimates to be smaller rather than larger as they are here.

There is a further issue that the estimates are inconsistent for small T large N; this can be examined using the Arrelano Bond estimator. There is also an issue as to whether the logarithmic is the best functional form, which needs to be examined.

5 Stock Adjustment

Given the import quantity series, S_{it}; we can calculate the stock of imported arms as

$$K_{it} = S_{it} + (1 ¡ ±)K_{it¡1}$$

given a depreciation rate ± and an estimate of initial stock K_{i0}; both of which can be estimated or assumed: For instance ± is likely to have a value of around 0:05 or 0:1. Then if $s_{it} = a_i + b_i t$; then $K_{i0} = a_i/±$: Stocks are likely to be a much smoother series than imports.

For non-zero stocks, we can determine a long-run level of stock as a function of say military expenditure,

$$K^{¤}_{it} = A_i M'_{it}$$

by a ...xed e¤ect levels regression of k_{it} on m_{it} which would give the long-run relationship (similar to a cointegrating relationship), other variables such as price and income could be added. Weapons are lumpy, imported in discrete units, e.g. squadrons of planes, battleships, etc. Thus we could model the short run adjustment function as

$$S_{it} = f(K^{¤}_{it}; K_{it})$$

where one would expect f() to be a non-linear function, e.g. no adjustment till $f(K^{¤}_{it}; K_{it}) > \cdot$ some threshold, then a jump to bring stocks up into line.

If we assume that no imports are made until the di¤erence between desired stock and actual stock is greater than a certain threshold $¹_i$; then the threshold amount plus some proportion, $0 < ¸ < 1$ of the di¤erence is made up. Thus some of our observations are corner solutions at zero. This would give a dynamic panel Tobit type relationship. One possible adjustment process is

$$\begin{aligned} S^{¤}_{it} &= ¸(K^{¤}_{it} ¡ K_{it¡1}) + "_{it}; \\ S_{it} &= S^{¤}_{it}; \quad \text{if } W^{¤}_{it} ¡ ¹_i > 0 \\ S_{it} &= 0; \quad \text{if } W^{¤}_{it} ¡ ¹_i \cdot 0 \end{aligned} \tag{1}$$

Another possibility is

$$\begin{aligned} S_{it} &= \left(\frac{K^{¤}_{it}}{K_{it¡1}}\right)^{¸} e^{"_{it}}; \quad \text{if } \left(\frac{K^{¤}_{it}}{K_{it¡1}}\right) e^{"_{it}} > ¹_i \\ S_{it} &= 0; \quad \text{if } \left(\frac{K^{¤}_{it}}{K_{it¡1}}\right)^{¸} e^{"_{it}} \cdot ¹_i \end{aligned} \tag{2}$$

This would give a logarithmic equation for the non-zero observations but would need to estimate the two relations (the probability of the observation being non-zero and the value it takes given that it is non zero) separately. Wooldridge 16.8.2&3 discusses similar models.

Our initial research will ...rst look at estimating such non-linear models on the non-zero sample and then extend it to the zero data sample, using a two stage estimator. This requires further discussion of the zeros.

6 Treatment of Zero and Missing Observations

The analysis above only used a subset of the full data available. We also consider a larger sample of 151 countries where there were more than 10

or more non-zero observed values during the period 1981-1999. This gives us 2869 observations. The observations can be a number, N; a zero, Z; or missing, M for each of SIPRI, S; or WMEAT, W. The joint and marginal distributions are

		S			
		N	Z	M	
	N	1678	15	453	2146
W	Z	117	12	480	609
	M	26	6	82	114
		1821	33	1015	2869

So in this sample of 151 countries we have 1678 observations where both WMEAT and SIPRI both record a number; in total SIPRI records 1821 numbers, ACDA 2146. Similarly for the other categories. Since WMEAT covers all arms transfers and SIPRI just major weapons systems, we should not observe cases where WMEAT records a zero and SIPRI a number, but we do, in 117 cases.

To examine the pattern of observations we can construct a two stage process. First for all observations, construct an indicator $y_{it} = 1$ if there is a number and $y_{it} = 0$ otherwise, then conditional on some explanatory variables (military expenditure, previous stocks, income, etc.) model

$$\begin{aligned} y^{\ast}_{it} &= \beta_0 x_{it} + u_{it} \\ y_{it} &= 0; \quad y^{\ast}_{it} \leq 0 \\ y_{it} &= 1; \quad y^{\ast}_{it} > 0 \end{aligned}$$

It would have to be done separately for the two measures as part of a system. Similar types of equations, where we had observations on both numbers, could be used directly to explain the decision to import or not, which would be the first stage of the non-linear logarithmic stock adjustment above.

There is an issue as to whether there is information is the pattern of zeros/missing observations. To examine this, for the observations where there is not a number, define $z_{it} = 1$ if there is a zero and $z_{it} = 0$ if it is missing

$$\begin{aligned} z^{\ast}_{it} &= \beta_0 x_{it} + u_{it} \\ z_{it} &= 0; \quad z^{\ast}_{it} \leq 0 \\ z_{it} &= 1; \quad z^{\ast}_{it} > 0 \end{aligned}$$

again this would be done jointly for the two systems. There are other possibilities, for instance we could use ordered models, where $y_i = 0$; missing,

$y_i = 1$; zero, $y_i = 2$ a number. Depending on the results for this analysis there are a number of possibilities. A simple one is to treat missing as zero. This would however be misleading if there were systematic under-recording for certain countries. This analysis may give us some feel for the probability that a missing observation is zero or just missed imports as a function of the observations on military expenditure etc.

7 Imputation.

Given that in many cases we have an observation from one source, but not another, there are potential gains from imputation, inferring the missing observation from the other. In addition, one possible source of di¤erences in observations is timing. It is often unclear exactly when a transfer took place, and SIPRI and WMEAT may record them in di¤erent years. Thus some form of time-averaging may improve estimation and allow us to impute more observations. We discuss procedures for using this information below.

Both SIPRI and WMEAT provide data on arms imports, say the other variables are W_{it} imports of weapons, M_{it}; military expenditure, P_{it} the price of imported weapons, Y_{it} GDP, POP_{it} population, for country $i = 1; 2; :::; N$ in year $t = 1; 2; :::; T$. The SIPRI estimate, S_{it} is a volume measure, and will be used to measure W_{it}. The ACDA estimate, A_{it} is a value measure, the ratio is thus a price index. In practice there are likely to be timing problems so we will estimate price as

$$P_{it} = \frac{A_{it}}{S_{it}}: \tag{3}$$

when the denominator, S_{it} is not zero. Unfortunately, in the observed data sets, the SIPRI estimate, S_{it} has not only zeroes but also missing values. We need …rst a procedure for dealing with this partial nonresponse problem. One important method for nonresponse adjustment is imputation.

7.1 Estimation in the presence of nonresponse

Nonresponse errors are the most known of the " nonsampling errors". They are attributed to causes other than the limitation of the investigation to a sample only, rather than the entire population. For estimation under ideal

conditions, consider the finite population of N elements with T time periods $U \{1, \ldots, i, \ldots, N\}$ and $\{1, \ldots, t, \ldots, T\}$, called the target population. To estimate, for example the price mean

$$\bar{P} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} P_{it}}{NT}. \tag{4}$$

where P_{it} is the value of the study variable of the price total, $P = \sum_{i=1}^{N} \sum_{t=1}^{T} P_{it}$.

Assume that s is a probability sample of size n, drawn from the target population U with probability $p(s)$. The inclusion probabilities, known for all $i, t \in U$, are than $\pi_{it} = \sum_{s \ni i,t} p(s)$. We assume that the design is such that $\pi_{it} > 0$ for all elements. Let $d_{it} = 1/\pi_{it}$ denote the design weight of element it. These weights are very important for computing estimators.

When the sampling design has been fixed, the inclusion probabilities π_{it} and the sampling design weights $d_{it} = 1/\pi_{it}$ are fixed, known quantities. Then the unbiased estimator of the price total P, is given by the Horvitz-Thompson estimator

$$\hat{P} = \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} P_{it}. \tag{5}$$

In the case of nonresponse, assume that response in the sample is obtained for the elements in a set denoted r with size m. Full response implies that $r = s$. Nonresponse implies that r is a proper set of s with size n. The nonresponse set is denoted $o = s - r$ with size $n - m$.

The imputation implies that proxy values are created for the values P_{it} that are missing because of nonresponse. The proxy value for the element $i, t \in o$, often called the imputed value for the it, is denoted $\hat{P}_{it}$. The completed data set will contain the same number of values as the originally intended sample s, that is, n values, and they are given by

$$P_{comp,it} = \begin{cases} P_{it} & \text{for } i, t \in r \\ \hat{P}_{it} & \text{for } i, t \in o \end{cases}. \tag{6}$$

The imputed Horvitz-Thompson estimator of the price total is of the form

$$\hat{P}_I = \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it} P_{comp,it}. \tag{7}$$

Below, we report two methods for obtaining completed data set. The first one, $\hat{P}_{comp1,it}$, imputes values using simple means of non-zero longitudinal observations, while the second one, $\hat{P}_{comp2,it}$, completes the data set

imputing values by using auxiliary information from nonmissing longitudinal observations of the other series.

7.1.1 Imputation by using simple means of non-zero longitudinal observations

We need prices for all i and t: A simple method to provide missing price observations is to estimate means over non zero estimates using longitudinal information of the data

$$\bar{P}_{:t} = \sum_{i=1}^{N} P_{it}/n_t; \tag{8}$$

where n_t is the number of non-zero observations in year t, and

$$\hat{D}_{i:} = \sum_{t=1}^{T} (P_{it} - \bar{P}_t)/n_i; \tag{9}$$

where n_i is the number of non-zero observations for country i, then

$$\hat{P}_{it} = \bar{P}_{:t} + \hat{D}_{i:}. \tag{10}$$

Using the estimated price information (3) and (10), we impute the missing values of prices as follows:

$$P_{comp1;it} = \begin{cases} P_{it} & \text{if } (A_{it} \neq 0 \text{ and } S_{it} \neq 0) \\ \hat{P}_{it} & \text{if } ((A_{it} \neq 0 \text{ and } S_{it} = 0) \text{ or } (A_{it} \neq 0 \text{ and } S_{it} = .) \\ & \quad \text{or } (A_{it} = . \text{ and } S_{it} \neq 0) \text{ or } (A_{it} = . \text{ and } S_{it} = .)) \\ 0 & \text{if } ((A_{it} = 0 \text{ and } S_{it} \neq 0) \text{ or } (A_{it} = 0 \text{ and } S_{it} = 0) \\ & \quad \text{or } (A_{it} = 0 \text{ and } S_{it} = .) \text{ or } (A_{it} = . \text{ and } S_{it} = 0)) \end{cases} \tag{11}$$

where $\hat{P}_{it}$ and 0s are the imputed values and $P_{comp1;it}$ show the completed price series.

7.1.2 Imputation by using auxiliary information from other series

In (11), the available information from one of the series is not used in the case of one of the series had missing value but the other one was observed. In order to utilise the available information in the imputation procedure, we employ the following price de…nition:

$$P_{it} = \frac{A_{it} + A_{it-1}}{S_{it} + S_{it-1}}. \tag{12}$$

If A_{it} is missing and S_{it} is non-zero then A_{it} may be estimated by

$$\tilde{A}_{it} = \hat{P}_{it}\, S_{it} = (\hat{P}_t^1 + \hat{D}_i^1)\, S_{it}; \tag{13}$$

and the price index imputed for A_{it} but observed for non-zero S_{it};

$$\hat{P}_{it}^{in} = \frac{\tilde{A}_{it} + A_{it-1}}{S_{it} + S_{it-1}}: \tag{14}$$

If A_{it} is non-zero and S_{it} is missing then S_{it} may be estimated by

$$\tilde{S}_{it} = \frac{A_{it}}{\hat{P}_{it}} = \frac{A_{it}}{(\hat{P}_t^1 + \hat{D}_i^1)}; \tag{15}$$

and the price index observed for non-zero A_{it} but imputed for S_{it};

$$\hat{P}_{it}^{ni} = \frac{A_{it} + A_{it-1}}{\tilde{S}_{it} + S_{it-1}}: \tag{16}$$

Using the estimated price information (3), (14) and (16), we obtain completed values of prices as follows:

$$P_{comp2;it} = \begin{cases} P_{it} & \text{if } (A_{it} \neq 0 \text{ and } S_{it} \neq 0) \\ \hat{P}_{it} & \text{if } ((A_{it} = 0 \text{ and } S_{it} \neq 0) \text{ or } (A_{it} = . \text{ and } S_{it} = .)) \\ \hat{P}_{it}^{in} & \text{if } (A_{it} = . \text{ and } S_{it} \neq 0) \\ \hat{P}_{it}^{ni} & \text{if } (A_{it} \neq 0 \text{ and } S_{it} = .) \\ 0 & \text{if } ((A_{it} \neq 0 \text{ and } S_{it} = 0) \text{ or } ((A_{it} = 0 \text{ and } S_{it} = 0) \\ & \text{or } ((A_{it} = 0 \text{ and } S_{it} = .) \text{ or } ((A_{it} = . \text{ and } S_{it} = 0)) \end{cases} : \tag{17}$$

where $\hat{P}_{it}$, $\hat{P}_{it}^{in}$, $\hat{P}_{it}^{ni}$, and 0s are imputed values and $P_{comp2;it}$ show completed price series.

To obtain Horvitz-Thompson estimators, all observations will be weighted with their own sampling design weights. In the ...rst imputation case, there are two types of inclusion probabilities; for observed (non-zero and zero) observations with π_{11}; and predicted observations with π_{12}: In the second imputation case, there are four types of inclusion probabilities; for observed (non-zero and zero) observations with π_{21}, for simply mean imputed using non-zero observations with π_{22}, and for two more imputed series using auxiliary information, π_{23}, and π_{24}.

8 Conclusion

Modelling Arms Imports raises some interesting technical questions. This is a preliminary progress report on a project to try and answer some of those questions. So far we have analysed some of the di¤erences in the properties of the two measures available and examined the relationship between them within a classical errors in variables framework. The patterns are broadly consistent with the SIPRI series being a volume measure and the WMEAT a value measure. Using a restricted panel, where there were no zero observations on either series, we have estimated simple demand functions for imports. These have sensible properties broadly consistent in terms of price response with earlier work. There do however seem to be interesting di¤erences between the cross-section dimension, where non-linearities seem important and the time-series dimension, where they do not. We have then discussed various possible directions for further research in terms of non-linear stock adjustment processes and the treatment of zero and missing observations. We would welcome comments on where we have got so far and where we hope to go.

References

Craft C and J.P. Smaldone (2002) The Arms Trade and the Incidence of Political Violence in Sub-Saharan Africa 1967-1997, Journal of Peace Research, 39,6, p693-710.

Levine P, F Mouzakis and R Smith (1998) Prices and Quantities in the Arms Trade, Defence and Peace Economics, 9 p223-236.

Levine P and R Smith (1997) The Arms Trade, Economic Policy 25,337-370.

Levine P and R Smith (2000) Arms Export Controls and Proliferation, Journal of Con‡ict Resolution,44,6, p885-895.

SIPRI Yearbook, Stockholm International Peace Research Institute, Armaments Disarmament and International Security, various years.

WMEAT, World Military Expenditures and Arms Transfers, US Department of State Bureau of Veri…cation and Compliance (previously Arms Control and Disarmament Agency), various years.

Wooldridge, J.M. (2002) Econometric Analysis of Cross Section and Panel Data, MIT Press.

Extended Sample.

1 AFGHANISTAN, 2 ALBANIA, 3 ALGERIA, 4 ANGOLA, 5 ARGENTINA,
6 ARMENIA 7 AUSTRALIA, 8 AUSTRIA, 9 AZERBAIJAN, 10 BAHRAIN
11 BANGLADESH, 12 BARBADOS, 13 BELARUS, 14 BELGIUM, 15 BELIZE
16 BENIN, 17 BHUTAN, 18 BOLIVIA, 19 BOSNA HERZIGOVA, 20 BOTSWANA
21 BRAZIL, 22 BRUNEI, 23 BULGARIA, 24 BURKINA FASO, 25 BURMA
26 BURUNDI, 27 CAMBODIA, 28 CAMEROON, 29 CANADA, 30 CAPE VERDE
31 CENTRAL AFRICAN REPUBLIC, 32 CHAD, 33 CHILE, 34 CHINA _MAINLAND, 35 CHINA _TAIWAN
36 COLOMBIA, 37 CONGO, 38 CONGO DEM. REP. Of KINSHASA, 39 CONGO REP. Of BRAZZEVILLE, 40 COSTA RICA
41 CROATIA, 42 CUBA, 43 CYPRUS, 44 CZECHOSLOVAKIA, 45 CZECH REPUBLIC
46 DENMARK, 47 DJIBOUTI, 48 DOMINICAN REPUBLIC, 49 ECUADOR, 50 EGYPT
51 EL SALVADOR, 52 EQUATORIAL GUINEA, 53 ERITREA, 54 ESTONIA, 55 ETHIOPIA
56 FIJI, 57 FINLAND, 58 FRANCE, 59 GABON, 60 GAMBIA, THE
61 GEORGIA, 62 GERMANY, EAST, 63 GERMANY, WEST, 64 GHANA, 65 GREECE
66 GUATEMALA, 67 GUINEA, 68 GUINEA-BISSAU, 69 GUYANA, 70 HAITI
71 HONDURAS, 72 HUNGARY, 73 ICELAND, 74 INDIA, 75 INDONESIA
76 IRAN, 77 IRAQ, 78 IRELAND, 79 ISRAEL, 80 ITALY
81 IVORY COAST, 82 JAMAICA, 83 JAPAN, 84 JORDAN, 85 KAZAKHSTAN
86 KENYA, 87 KOREA, NORTH, 88 KOREA, SOUTH, 89 KUWAIT, 90 KYRGYZSTAN
91 LAOS, 92 LATVIA, 93 LEBANON, 94 LESOTHO, 95 LIBERIA
96 LIBYA, 97 LITHUANIA, 98 LUXEMBOURG, 99 MACEDONIA, 100 MADAGASCAR
101 MALAWI, 102 MALAYSIA, 103 MALI, 104 MALTA, 105 MAURUTIANA
106 MAURITIUS, 107 MEXICO, 108 MOLDOVA, 109 MONGOLIA, 110 MOROCCO

111 MOZAMBIQUE, 112 NAMIBIA, 113 NEPAL, 114 NETHERLANDS, 115 NEW ZEALAND

116 NICARAGUA, 117 NIGER, 118 NIGERIA, 119 NORWAY, 120 OMAN

121 PAKISTAN, 122 PANAMA, 123 PAPUA NEW GUINEA, 124 PARAGUAY, 125 PERU

126 PHILIPPINES, 127 POLAND, 128 PORTUGAL, 129 QATAR, 130 ROMANIA

131 RUSSIA, 132 RWANDA, 133 SAO TOME & PRINCIPE, 134 SAUDI ARABIA, 135 SENEGAL

136 SERBIA AND MONTENEGRO, 137 SIERRA LEONE, 138 SINGAPORE, 139 SLOVAKIA, 140 SLOVENIA

141 SOMALIA, 142 SOUTH AFRICA, 143 SOVIET UNION, 144 SPAIN, 145 SRI LANKA

146 SUDAN, 147 SURINAME, 148 SWAZILAND, 149 SWEDEN, 150 SWITZERLAND

151 SYRIA, 152 TAJIKISTAN, 153 TANZANIA, 154 THAILAND, 155 TOGO

156 TRINIDAD & TOBAGO, 157 TUNISIA, 158 TURKEY, 159 TURKMENISTAN, 160 UGANDA

161 UKRAINE, 162 UNITED ARAB EMIRATES, 163 UNITED KINGDOM, 164 UNITED STATES,165 URUGUAY

166 UZBEKISTAN, 167 VENEZUELA, 168 VIETNAM, 169 YEMEN (ADEN), 170 YEMEN (SANAA)

171 YUGOSLAVIA, 172 ZAIRE, 173 ZAMBIA, 174 ZIMBABWE,