Efficient Management of Multi-Frequency Panel Data with Stata

Christopher F Baum

Department of Economics, Boston College

May 2001

Prepared for United Kingdom Stata User Group Meeting

`http://repec.org/usug01/baum.UKSUG.pdf`

Empirical research in macroeconomics and international finance increasingly depends on panel data, and econometric techniques that make use of those data (e.g. panel unit-root tests)

Panel data for macroeconomic series may be available at different timeseries frequencies. For instance, national income accounts data (GDP, consumption, investment) are available at no more than quarterly frequencies. Price indices and many financial data (money supply, international reserves, etc.) are generally available at no more than monthly frequency. Financial markets' prices and volumes are readily available daily, and may be accessible at higher frequencies.

This paper describes a research project in international finance which makes use of both monthly-frequency economic data and daily-frequency financial market data in order to generate a more effective measure of volatility at the monthly frequency. The motivation for this process was provided by Robert Merton (1980) in the context of stock market volatility. Following Klaassen (1999), we will apply it to the analysis of exchange rate volatility, so that we will generate a monthly series for volatility based upon the intra-month—daily—variations in the exchange rate. Contrast this approach with an (G)ARCH model on monthly data, or an (exponentially) weighted moving average of recent monthly data. Both have been shown to possess deficiencies in modelling the stylized facts of exchange rate uncertainty.

In this project, we specify and estimate a model of bilateral trade flows in an n-country world. Unlike much of the empirical literature, we do not focus upon the USA as the "home country." We consider the real exports of each of the $n=18$ countries to each of the other countries, for a total of 306 potential models of trade. The analytical structure we employ distinguishes between the exporter and importer, so that each distinct trade flow is of interest.

From a data management perspective, this implies that we have an unusual sort of panel: not of 18 countries over T time periods, but of 306 i, j pairs over T time periods. This implies that measures of exchange rate volatility will depend upon both the exporting and importing country, and the spot exchange rate between those countries.

The equation to be estimated for real exports (x_{it}) is

$$x_{it} = x\left(y_{jt}, E_{t-\tau}\left[s_t\right], \sigma_{s,t-\tau}\left[s_t\right], \sigma_{s,t-\tau}\left[s_t\right] \times \sigma_{y,t-\tau}\left[y_t\right]\right)$$
$$(1)$$

where y_{jt} is foreign real income, $E_{t-\tau}\left[s_t\right]$ is the expectation of the spot exchange rate formed τ periods ago, and $\sigma_{s,t-\tau}\left[s_t\right]$ is its volatility estimated τ periods ago. The last term captures the interaction between exchange rate volatility and foreign income volatility, $\sigma_{y,t-\tau}\left[y_t\right]$.

We now consider how these measures can be constructed and assembled for estimation of this equation for each i, j pair.

The nominal export data were obtained from the IMF's *Direction of Trade Statistics*, which reports monthly trade flows between every country pair. These data, as retrieved from an SQL database, were in the "long long" format, and were converted to "long wide" via Stata's `reshape` facility, in which each country's exports are recorded in one xt-format variable, with identifier $ccode$ denoting the importing partner country, and identifier $month$ specifying the period of the flow. This is then a balanced panel with missing data values appearing for the "own" flows. See Figures 1 and 2.

| "i" | "j" | | |
Exporter	Importer	Month	Real Exports
USA	CAN	1980m1	nnn.n
USA	CAN	1980m2	nnn.n
USA	CAN	1980m3	nnn.n
...	...	...	...
USA	CAN	1998m12	nnn.n
USA	UK	1980m1	nnn.n
USA	UK	1980m2	nnn.n
USA	UK	1980m3	nnn.n
...	...	...	...
USA	UK	1998m12	nnn.n
	. . .		
	. . .		
	. . .		
USA	CH	1980m1	nnn.n
USA	CH	1980m2	nnn.n
USA	CH	1980m3	nnn.n
...	...	...	...
USA	CH	1998m12	nnn.n
CAN	USA	1980m1	nnn.n
CAN	USA	1980m2	nnn.n
CAN	USA	1980m3	nnn.n
...	...	...	...
CAN	USA	1998m12	nnn.n
	. . .		
	. . .		
	. . .		
CH	SE	1980m1	nnn.n
CH	SE	1980m2	nnn.n
CH	SE	1980m3	nnn.n
...	...	...	...
CH	SE	1998m12	nnn.n

Figure 1:
"Long Long" format

as retrieved from
IMF DOTS

228 months per
country-pair for
each of the 306
country-pairs

69,768 observations

ccode	month	exp111	exp112	exp122	. . .	exp548
111	1980m1		nnn.n	nnn.n		nnn.n
111	1980m2	(missing)	nnn.n	nnn.n		nnn.n
. . .	...		...	...		...
111	1998m12		nnn.n	nnn.n		nnn.n
112	1980m1	nnn.n		nnn.n		nnn.n
112	1980m2	nnn.n	(missing)	nnn.n		nnn.n
. . .	...	...		...		...
112	1998m12	nnn.n		nnn.n		nnn.n
122	1980m1	nnn.n	nnn.n			nnn.n
122	1980m2	nnn.n	nnn.n			nnn.n
. . .	...	...	...	(missing)		nnn.n
122	1998m12	nnn.n	nnn.n			...
. . .						nnn.n
. . .						
548	1980m1	nnn.n	nnn.n	nnn.n		
548	1980m2	nnn.n	nnn.n	nnn.n		(missing)
. . .	...	...	...	...		
548	1998m12	nnn.n	nnn.n	nnn.n		

Figure 2: "Long Wide" format

Reshaped into conventional panel (xt) format

228 months per exporting country for each of the 18 countries (3,876 usable observations per exporter; exports to own country treated as missing)

Variable exp111 for ccode 112: exports of country 111 (USA) to country 112 (UK).
Variable exp112 for ccode 111: exports of country 112 (UK) to country 111 (USA), i.e. US imports from UK, but measured on the basis of reported exports, not that of US imports (which would differ by the amount of statistical discrepancy (SD) in the estimates).

In the past the SD in the trade accounts has approached the magnitude of the UK trade deficit!

The estimated model makes use of other series that are specific to the country but not to the relationship (partner country), such as national income. Those series were retrieved from the IMF's *International Financial Statistics* database. For a given exporting country i, the model contains real foreign income of the importing country j, so that the jth income series must be matched with that country's identifiers throughout the dataset. Stata's `merge` facility makes that straightforward, supporting the "one-to-many" replication of country j's income timeseries, aligned properly with each instance in which that country is the importer.

A similar task is required to generate an estimated volatility series for real income. Since we do not have higher-frequency data for income, this series must be generated at the monthly frequency, using (log) industrial production as a proxy. The technique employed by Thursby and Thursby (1987) is used: log IP is regressed on a quadratic trend for a 12-month moving window, and the RMSE is used as the next period's estimate of income volatility. Although Stata does not have explicit commands for such a "rolling regression," the procedure was readily programmed (in Stata 6) making heavy use of Nick Cox' `listutil` functions. It would be even easier to implement in Stata 7, with `foreach` available. The `merge` technique used above is then used to match the generated income volatility series − which is country-specific, but not relationship-specific − with the appropriate blocks in the "long wide" panel data set.

The monthly measure of exchange rate volatility proposed by Merton (1980) aggregates (squared) intra-monthly changes in the exchange rate in order to capture that month's volatility. We focus on real exchange rates: that is, the spot (nominal) exchange rate adjusted for relative prices in the two countries. Real exchange rate volatility will therefore depend upon not only the variability in the foreign exchange market but also on the volatility of prices. The spot exchange rate is available at a business-daily frequency, but the price indices are not. We linearly interpolate the relative price over the month's business days to generate a calendar-daily real exchange rate series (expressed in log form).

The squared first difference of the daily log real exchange rate $\left(s_t^d\right)$, after dividing by the square root of the number of days intervening, is then defined as the daily contribution to monthly volatility:

$$\varsigma_t^d = \left(100\frac{\Delta s_t^d}{\sqrt{\Delta \tau_t}}\right)^2, \qquad (2)$$

where the denominator expresses the effect of calendar time elapsing between observations on the s process. If data were available every calendar day, $\Delta \tau_t = 1, \forall t$, but given that exchange rate data are not available on weekends and holidays, $\Delta \tau_t \in (1, 5)$.

The estimated monthly volatility of the (log) real exchange series is defined as

$$\Phi_t\left[s_t\right] = \sqrt{\sum_{t=1}^{T} \varsigma_t^d} \qquad (3)$$

where the time index for $\sigma_t\left[s_t\right]$ is at the monthly frequency.

The construction of the daily measure cannot be performed under the time series calendar, since the formula for exchange rate changes allows for missing values attributable to weekends, holidays, etc. Thus the change between the spot rate and the last available quotation must be calculated, and the interval of calendar time between those dates used to properly scale the observed change. Once the series of scaled changes is computed, it is squared and cumulated over the month using the `sum` function in order to calculate $\Phi_t\left[s_t\right]$. That series may then be `merged` back onto the monthly dataset after discarding all but the end-of-month observations.

With all data series required in the estimation constructed, the model of real exports given in (1) may be specified, following Klaassen (1999), as an infinite distributed lag in its determinants. The sequences of coefficients on the lag terms must be constrained to render the model estimable. Restrictive specifications–such as those imposing monotonicity, linearity, or exponential decay upon the lag coefficients–may be quite harmful, and inadequate dynamics embedded in the model's specification will almost surely result in damaging omitted-variable bias. We employ the parsimonious specification proposed by Klaassen, consisting of a Poisson lag:

$$\beta_{k\tau} = \beta_k^* \left[\frac{(\lambda_k - 1)^{\tau-1}}{(\tau - 1)!} e^{-(\lambda_k - 1)} \right], \qquad (4)$$

where $\lambda_k > 1$ and k indexes the explanatory variables, each of which is associated with a β^* parameter.

This lag structure encompasses several more restrictive alternatives, such as the geometric lag. The parameter λ_k, as Klaassen points out, is approximately the mode of the (translated) Poisson distribution, and is the lag at which the maximal effect of the regressor occurs. In the estimation, constraints are imposed on the vector of λ parameters related to the exchange rate so that the same λ is used for the expected real exchange rate $E_{t-\tau}[s_t]$, its volatility $\sigma_{s,t-\tau}[s_t]$, and the interaction term $\sigma_{s,t-\tau}[s_t] \times \sigma_{y,t-\tau}[y_t]$.

This approach allows us to parsimoniously capture declining as well as hump-shaped lag structures, and permits the mean lag length of the distributed lag to be estimated endogenously rather than imposed on the data.

A tradeoff exists between the length of lag allowed in the Poisson specification and the sample size over which the model is fit; we found that allowing up to $L =30$ months' lag generated sensible results in almost all cases. The bracketed term in equation (4) can be considered the weight placed on period $t - \tau$'s value of the regressor: ϖ_τ, where $\varpi_\tau > 0$ and $\sum_{\tau=1}^{\infty} \varpi_\tau = 1$. The latter constraint is not imposed in the estimation–as it is a constraint on the infinite sequence of Poisson lag coefficients, not the finite subset utilised in the model–but may be evaluated from the estimated parameters, in terms of the value of $\sum_{\tau=1}^{L} \widehat{\varpi}_\tau$.

Estimation of the model of real exports is performed with Stata's nonlinear least squares (`nl`) algorithm.

A major concern, given the hundreds of models to be estimated and their results tabulated, is the efficient handling of this process within Stata. This goal has been achieved by making use of several useful tools:

- Various `listutil` (Cox, 2000) functions to set up the models to be estimated: every country's real exports vs every trading partner's variables over the appropriate estimation sample (one timeseries slice of the panel)

- `testnl` to transform the estimated coefficients into the statistics of interest from the modelling perspective

- `postfile` to capture the key results from each model and assemble them in a separate dataset for further statistical and graphical analysis

Fragment of code employing `listutil`:

```
local cty 112 122 128 132 134 136 138 142 144 /*
        */ 146 156 158 172 182 184 534 548 111
wclist 'cty'
local nvs = 1- 'r(nw)'
local i 1

while 'i' <= 18 {
        rotlist 'cty',rot(-'i')
        local now 'r(list)'
        takelist 'now',take(1)
        local dep 'r(list)'
        takelist 'now',take('nvs')
        local mod 'r(list)'
        di " "
        di " dep : 'dep'  vs 'mod'"
        di " "
        for any 'mod': nlpoist6 30 lnipX /*
     */ lnsrX ssqX lnrxSX volX 'dep' X 250
        local i = 'i'+1
        }
```

The coefficients estimated by `nl` are not themselves the λ terms of the model, but transformations of those terms. Tests on the λs themselves are constructed with `testnl`. Certain effects in the model are nonlinear, due to the presence of the interaction term of exchange rate volatility and foreign income volatility. These effects are calculated via Stata's `lincom`, which constructs point and interval estimates of the derivatives of interest. For instance,

$$\frac{\partial x_{ijt}}{\partial \sigma_{ijs,t} \left[s_{ijt} \right]} = \widehat{\beta}_3 + \widehat{\beta}_4 \sigma_{jy,t} \left[y_{jt} \right], \qquad (5)$$

so that, given $\widehat{\beta}_4 \neq 0$, the effect of real exchange rate volatility depends on the level of foreign income volatility, with the sign of the interaction being that of $\widehat{\beta}_4$. That is, greater foreign income volatility could either enhance or diminish the direct effect of exchange rate volatility, depending on the signs of the two estimated coefficients.

Given nearly 300 estimated models, scrutiny of the individual models' results is not workable. We utilize `postfile` within the model estimation loop to create a dataset with all relevant elements of each model's performance, identified by the country code and partner country code. The contents of the resulting dataset render analysis of the distribution of key effects − e.g. the estimated impact of exchange rate volatility on export flows − quite straightforward. We can readily analyze any patterns that may appear in the magnitude or significance of these effects, by exporter or by importer, to gain a better perspective on the model's success.

An example of one of the graphs produced to analyze these findings is provided in Figure 8 (of the working paper). Perhaps someday we will be able to easily produce this sort of graphic within Stata!

Figure 8
Estimated Sensitivity to Exchange Rate Volatility
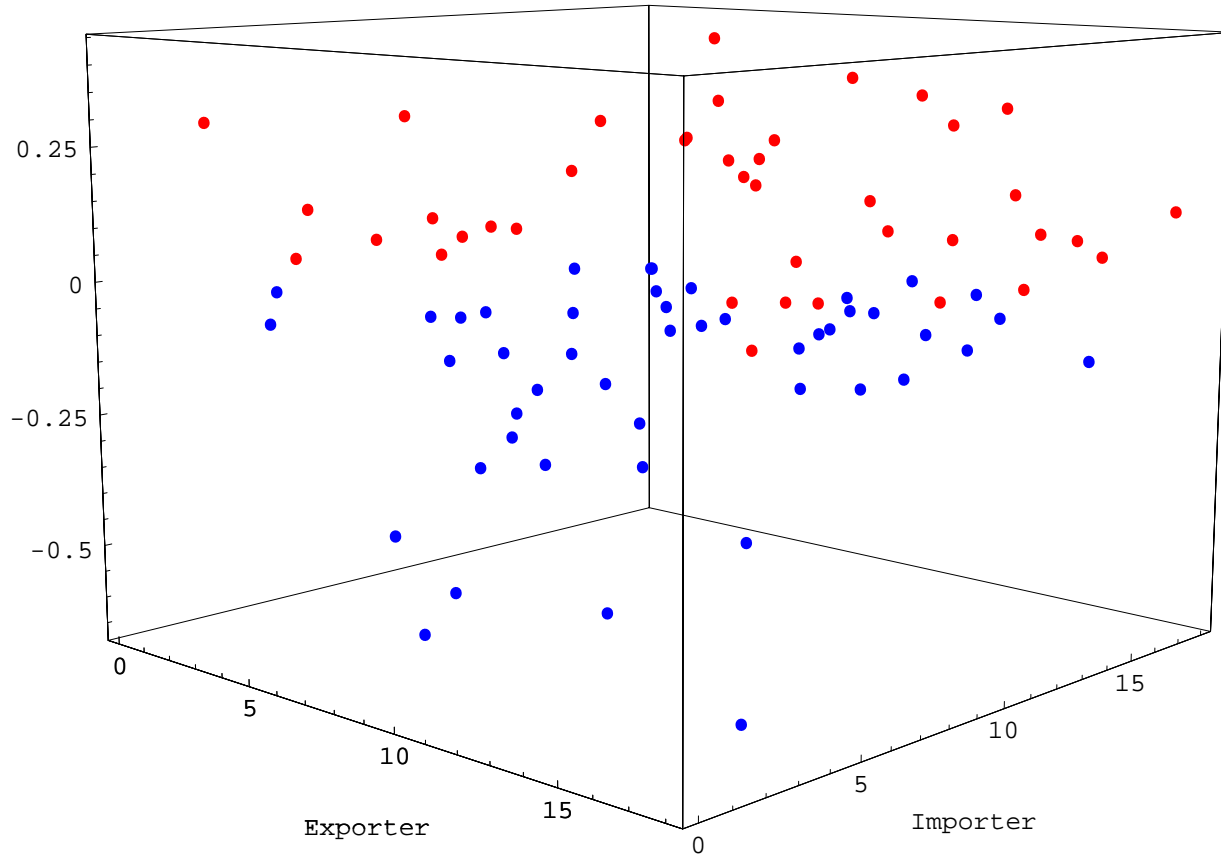
Copies of the Stata programs used in this research are available on request. For the most part, they are quite specific to the form of the model and dataset employed.

For those interested in the subject matter of this research and our findings, please see "Exchange Rate Effects on the Volume of Trade Flows: An Empirical Analysis Employing High-Frequency Data," C F Baum, Mustafa Caglayan and Neslihan Ozkan, Boston College Economics Working Paper No. 488. Available from IDEAS.

Works cited:

Klaassen, F. (1999), Why is it so difficult to find an effect of exchange rate risk on trade? Unpublished working paper, CentER.

Merton, Robert (1980), On estimating the expected return to the market: An exploratory investigation. Journal of Financial Economics, 8, 323-361.