Online Appendix

Exchange Rate Disconnect and Private Information: What Can we Learn from Euro-Dollar Tweets?

Vahid Gholompour and Eric van Wincoop

August 2018

This online Appendix discusses the computation of model moments. We first characterize the distribution of TS_t^i conditional on the average expectation. We then discuss approach to computing moments other than correlations. We finally discuss correlation moments.

I Distribution TS_t^i

For what follows, it is useful to characterize the distribution of TS_t^i conditional on the average expectation, which we will denote $x_t = \bar{E}_t(s_{t+T} - s_t)$.

Then $E_t^i(s_{t+T} - s_t) = x_t + \mathbf{z}'_T\mathbf{M}\mathbf{w}_t^i$. Let σ_w be the standard deviation of the second term, associated with signal errors. We can write

$$TS_t^i = TS_t(x_t) + \epsilon_t^i \tag{1}$$

Here $TS_t(x_t)$ is the mean value of TS_t^i conditional on x_t. This is equal to the average Twitter Sentiment if there were an infinite number of tweets that day. We have

$$TS_t(x_t) = 1 - \Phi\left(\frac{c - x_t}{\sigma_w}\right) - \Phi\left(\frac{-c - x_t}{\sigma_w}\right) \tag{2}$$

where $\Phi(.)$ is the cumulative normal distribution. It follows that

$$\epsilon_t^i = \begin{cases} 1 - TS_t(x_t) & \text{with probability} \quad 1 - \Phi\left(\frac{c-x_t}{\sigma_w}\right) \\ -1 - TS_t(x_t) & \text{with probability} \quad \Phi\left(\frac{-c-x_t}{\sigma_w}\right) \\ -TS_t(x_t) & \text{with probability} \quad \Phi\left(\frac{c-x_t}{\sigma_w}\right) - \Phi\left(\frac{-c-x_t}{\sigma_w}\right) \end{cases} \tag{3}$$

We now know the distribution of Twitter Sentiment of individual agents conditional on x_t. Below we will use in particular the variance $var(\epsilon_t^i)$ conditional on x_t.

II Approach to Computing Model Moments

In order to estimate the model parameters, discussed in Section 4.2 of the paper, we need to compute the model moments. We focus on the 11 moments listed in

Table 3. In principle the model moments correspond to the average across an infinite number of simulations of the model over the 1103 days for which we have data. In practice model moments are usually computed as the average over a finite number of simulations, like 1000. When considering different sets of model parameters, the model moments are computed using the same set of shocks for the simulations. In our case the shocks are $\varepsilon_t^f, \varepsilon_t^b$ and $\epsilon_t^{v,i}$. However, 1000 simulations, or any other reasonably finite number, creates too much inaccuracy in the context of our application. The reason is that Twitter Sentiment is a discrete variable, so that for a given set of shocks a tiny change in model parameters can lead to a discrete change in TS_t^i for some days and agents, which leads to a discrete change in various moments. Such discontinuities create problems in estimating the parameters as moments are not smooth functions of parameters.

We resolve this as follows. Realizations of the signal error shocks $\varepsilon_t^{v,i}$ translate into realizations of ϵ_t^i, whose distribution is given by (3). We can then write a specific sample moment as $m = m(\mathbf{e}, \mathbf{x})$, where $\mathbf{e}$ consists of the realizations of ϵ_t^i and $\mathbf{x}$ consists of the realizations of the shocks ε_t^f and ε_t^b. We need to compute the mean of $m(\mathbf{e}, \mathbf{x})$ across all possible outcomes for $\mathbf{e}$ and $\mathbf{x}$. We do so by first computing a theoretical expression for the mean across all possible outcomes for $\mathbf{e}$. This theoretical expression is for one particular set of values of $\mathbf{x}$. We next simulate the model 1000 times by drawing the shocks ε_t^f and ε_t^b in order to approximate the mean of the moment across all values of $\mathbf{x}$.

We double check that the model moments obtained this way are the same as obtained by simulating across all shocks, including the $\varepsilon_t^{v,i}$. We have done this for 100,000 simulations for a particular parameterization. While it is possible to do this for one set of parameters, it is extremely time-consuming (it takes 8 hours) and therefore runs into computational constraints when estimating parameters. In addition, even for such a large number of simulations the moments are still not completely smooth functions of the parameters when simulating across all shocks, including the $\varepsilon_t^{v,i}$.

III Moments other than Correlations

First consider the sample variance of Twitter sentiment for group j. We can write

$$TS_t = TS_t(x_t) + \frac{1}{n_t}\sum_{i=1}^{n_t}\epsilon_t^i \tag{4}$$

Let S stand for the number of days in the sample (here 1103) as well as the set of days in the sample. Then the sample variance is equal to

$$\frac{1}{S-1}\sum_{t\in S}\left(TS_t(x_t) + \frac{1}{n_t}\sum_{i=1}^{n_t}\epsilon_t^i\right)^2 - \frac{S}{S-1}\left[\frac{1}{S}\sum_{t\in S}\left(TS_t(x_t) + \frac{1}{n_t}\sum_{i=1}^{n_t}\epsilon_t^i\right)\right]^2$$

In this case $\mathbf{x}$ consists of the values of x_t in the sample, which only depend on the fundamental and noise shocks. We first compute the theoretical mean of this expression for given values of x_t over the distribution of the ϵ_t^i given in (3). Doing so gives

$$var(TS_t(x_t)) + \sum_{t\in S}\frac{1}{Sn_t}var(\epsilon_t^i) \tag{5}$$

Here the first variance is the sample variance of $TS_t(x_t)$, while $var(\epsilon_t^i)$ is computed from the distribution (3) for given x_t. We then finally take the mean across realizations of x_t across 1000 simulations of the model. When simulating the model over 1103 days, we always assume that the number of tweets each day, n_t, corresponds exactly to that in the data for that day.

Next consider the disagreement moments. It is measured as the average cross-sectional variance of Twitter Sentiment across the days in the sample, which is equal to

$$\frac{1}{S}\sum_{t\in S}\frac{n_t}{n_t-1}\left(\frac{1}{n_t}\sum_{i=1}^{n_t}\left(\epsilon_t^i\right)^2 - \left(\frac{1}{n_t}\sum_{i=1}^{n_t}\epsilon_t^i\right)^2\right) \tag{6}$$

The mean across the distribution of ϵ_t^i is

$$\frac{1}{S}\sum_{t\in S}var(\epsilon_t^i) \tag{7}$$

where the variance is again computed from (3) as a function of x_t. We finally take the mean across 1000 simulations of the model that lead to different values of x_t.

Next consider the directional moments based on a subsequent change in the exchange rate over the next m days. The sample moment is equal to

$$\frac{1}{\sum_{t \in S} n_t} \sum_{t \in S} \sum_{i=1}^{n_t} u_t^i \tag{8}$$

where

$$u_t^i = \begin{cases} 1 & \text{if} \quad \text{sign}(TS_t^i) = \text{sign}(s_{t+m} - s_t) \\ -1 & \text{if} \quad \text{sign}(TS_t^i) = -\text{sign}(s_{t+m} - s_t) \\ 0 & \text{if} \quad TS_t^i = 0 \end{cases} \tag{9}$$

The theoretical mean of the sample moment across realizations of ϵ_t^i is

$$\frac{1}{\sum_{t \in S} n_t} \sum_{t \in S} n_t TS_t(x_t)\text{sign}(s_{t+m} - s_t) \tag{10}$$

We again take the average across $\mathbf{x}$ through 1000 simulations. Note that the exchange rate change is part of $\mathbf{x}$ as it depends on the shocks ε_t^f and ε_t^b.

IV Correlation Moments

Here we consider the remaining moments: the correlation between Twitter Sentiment of the informed and uninformed and the moments involving correlations between Twitter Sentiment and exchange rate changes.

First a general comment is in order about all correlation moments. In principle we should first write down an expression for the sample correlation and then take the mean over realizations of ϵ_t^i. However, the resulting expression is a complicated non-linear function of the ϵ_t^i for which we cannot compute the theoretical mean. We therefore proceed slightly differently. We first compute the mean over realizations of ϵ_t^i for the covariance and the variance of both variables. We then use this result to compute the correlation (covariance divided by product of the standard deviations). It turns out that this delivers virtually identical results. We have verified this by numerically computing the correlation both ways for given realizations of ε_t^f and ε_t^b. In addition, we have compared the results based on 100,000 simulations of all the shocks over 1103 days $(\varepsilon_t^b, \varepsilon_t^f, \epsilon_t^{v,i})$, computing the correlation as the average across the simulations, to the moments obtained with our approach. The results are again virtually identical. The reason it makes little difference is that across different draws for the ϵ_t^i across simulations, the variance of variables entering correlations changes very little.

Consider predictive correlations. These are the correlation between the Twitter Sentiment Index and the change in the exchange rate over the next m days. We first consider the sample covariance between TS_t and $s_{t+m} - s_t$. Since $s_{t+m} - s_t$ only depends on realizations of ε_t^f and ε_t^b, the average of this sample covariance across realizations of ϵ_t^i is equal to the sample covariance between $TS_t(x_t)$ and $s_{t+m} - s_t$. This is because $TS_t(x_t)$ is the mean of TS_t across realizations of ϵ_t^i. We then divide the sample covariance between $TS_t(x_t)$ and $s_{t+m} - s_t$ by the product of the square root of the variance of TS_t and $s_{t+m} - s_t$. We discussed the variance of TS_t in Section III. The variance of $s_{t+m} - s_t$ only depends on realizations of ε_t^f and ε_t^b. We finally again take the mean of the resulting correlation across values of ε_t^f and ε_t^b over 1000 simulations.

The final moment involving the relationship between Twitter Sentiment and exchange rate changes is the contemporaneous weekly correlation between Twitter Sentiment and the exchange rate change. We only consider non-overlapping weeks. Let w be a particular week (5 days in the model). The average Twitter Sentiment in week w is

$$\frac{1}{5}\sum_{m=1}^{5} TS_{5*(w-1)+m} \tag{11}$$

The exchange rate change in the corresponding week is s_{5w}-s_{5w-4}. Since again the exchange rate change does not depend on the ϵ_t^i, the mean of the sample covariance is equal to the sample covariance of

$$\frac{1}{5}\sum_{m=1}^{5} TS_{5*(w-1)+m}(x_{5*(w-1)+m}) \tag{12}$$

and s_{5w}-s_{5w-4}.

In order to compute the correlation we also need the variance of (11). This can be computed analogous to the variance of TS_t. After writing down the sample variance and taking the mean across the distribution of the ϵ_t^i, we get

$$var\left(\frac{1}{5}\sum_{m=1}^{5} TS_{5*(w-1)+m}(x_{5*(w-1)+m})\right) + \sum_{t\in S}\frac{1}{5Sn_t}var(\epsilon_t^i) \tag{13}$$

where S is the number of days in the sample.