AGGREGATION & MODEL-BASED METHODS IN SEASONAL ADJUSTMENT OF LABOR FORCE SERIES

Stuart Scott and Peter Zadrozny, U.S. Bureau of Labor Statistics
Stuart Scott, 2 Massachusetts Avenue, NE, room 4915, Washington, DC 20212

Keywords: X-12-ARIMA, TRAMO/SEATS, direct and indirect adjustment, unemployment rate

1. Introduction

One of the most closely-watched numbers produced by the U.S. Bureau of Labor Statistics (BLS) is the seasonally adjusted civilian unemployment rate. It is computed from eight employment and four unemployment series. These and other labor force statistics are derived from the Current Population Survey (CPS), a survey of close to 50,000 households. A motivation for examining the aggregation issue is that four of twelve components of the unemployment rate are for teenagers and four are agricultural employment. With the decline in size of the farm sector and reductions in the CPS sample, it makes sense to examine adjustment quality. This study revisits work by Estela Dagum (1978) for the Levitan Commission.

An impetus for the use of models is the development by Augustin Maravall and Victor Gomez (1997) of the TRAMO/SEATS software for adjustment based on seasonal ARIMA models. TRAMO/SEATS offers advances in model identification and signal extraction, making the use of models more practical, and its availability makes these methods more accessible to the public. Fischer and Planas (1998) for Eurostat have worked extensively with the package.

2. Aggregation

As practiced, seasonal adjustment is nonlinear, i.e., if A denotes the seasonal adjustment filter, $A(X+Y) \neq A(X) + A(Y)$. X-11 starts with linear filters, so a "pure" adjustment with its filters is linear. Several factors can introduce nonlinearity: X-11 robustness features, ARIMA extrapolation, estimation of trading day or other calendar effects, differing modes of seasonal adjustment (additive, multiplicative). In addition, our most important statistic, the unemployment rate, is a ratio.

Even small departures from linearity can cause problems. There's a story that one month major components of the old Wholesale Price Index went up, while the total went down. When Julius Shiskin, then BLS commissioner, appeared before Congress, he was asked whether or not the Bureau knew how to do arithmetic. This led Shiskin to expand the use of indirect adjustment, which provides for consistency. With indirect adjustment (IA), the seasonally adjusted aggregate is computed from seasonally adjusted components (thus insuring consistency); with direct adjustment (DA), the aggregate itself is seasonally adjusted. For labor force series, there are so many series of interest, based on demographics, geography, job characteristic (e.g., industry, full-time vs. part-time), that it is not feasible to enforce complete consistency across all published series. Currently, the four series determined by gender and age (teenage, adult) are seasonally adjusted directly for unemployment, agricultural employment, and nonagricultural employment, and all aggregate series derived from these 12 series, including the civilian unemployment rate, are seasonally adjusted indirectly. This system is satisfying, because the seasonally adjusted data are consistent for all these series, among the most important products of the CPS. Table 1 shows codes used in the paper for the 12 basic components and a few aggregates. (The first character denotes age; the second, gender; and the last two, labor force status).

The question that arises is whether this use of IA sacrifices quality. This is largely an empirical question; in principle, either method could be better, depending on the application. Let's think about the case of only two components.

Case 1. Components are seasonal, but "noisy."
Often the aggregate will be less noisy, so
DA may be better.

Case 2. Components are highly seasonal, but with different patterns.

Table 1. Series codes

YFUN	Unemployment, teenage females
YMUN	Unemployment, teenage males
AFUN	Unemployment, adult females
AMUN	Unemployment, adult males
YFEA	Agricultural employment, teenage females
YMEA	Agricultural employment, teenage males
AFEA	Agricultural employment, adult females
AMEA	Agricultural employment, adult males
YFEN	Nonag employment, teenage females
YMEN	Nonag employment, teenage males
AFEN	Nonag employment, adult females
AMEN	Nonag employment, adult males
YUN	Unemployment, teenagers
FUN	Unemployment, females
MEA	Agricultural employment, males
AEN	Nonag employment, adults
etc.	

The patterns may "wash out" somewhat in the aggregate, so the combined pattern may be more difficult to estimate. Here, IA may be better.

Assuming one wishes to set up an indirect adjustment, Dagum (1978) gives four principles to consider in selecting components.

(1) identifiable seasonality,
(2) simple pattern,
(3) pattern distinguishable from other components,
(4) meaningful economic entity.

The first is obvious. (2) is desirable, since a complex pattern may be due to a lack of homogeneity of the series, i.e., it may be the aggregate of series with different patterns. If two potential components have a similar seasonal pattern, then, according to (3), it may make sense to combine them. Again, the last is fairly obvious: it would not be desirable to form a component from two series simply on the basis of a similar pattern, rather than their economic sense. Such a series would not be very usable. An in-depth discussion of these principles is contained in Dagum's paper.

Dagum recommended reducing the number of components for application of IA from 12 to 7. In agreement with Dagum, this paper will show that in some cases DA has better properties than IA, and that IA outperforms DA for the civilian unemployment rate. However, given the many aggregates that are derived from the basic set of 12 series, it will argue against combining any components in the current set.

A variety of statistics can be used to compare DA and IA; among them are the Shiskin-Dagum quality control statistics (M1-M11 and summary Q), sliding span statististics, spectra, and revision statistics (cf. Findley et al, 1998). In general, Q statistics tend to favor DA. An initial look at spectra of seasonally adjusted series fails to show appreciable differences. Some use is made of revision statistics. Tables 2 and 3 contain sliding span statistics, the principle evaluation tool. As discussed in Findley et al, for each month common to two or more sliding spans, a maximum per cent difference (MPD) is computed. It is desirable that most MPD's be acceptably small. We focus on the 85^{th} percentile of MPD's for seasonal factors or seasonally adjusted series, on the 60^{th} percentile for month-to-month change (change being more volatile), and on the maximum. Only nine years of data, 1990-98, are used, since pre-1990 data are not consistent with 1990 Census data. This has limited us to applying sliding spans with three 7-year spans.

Two of the smallest and noisiest series are YFEA and YMEA, agricultural employment for teenage females and males, respectively. Even so, the quality control statistics support seasonal adjustment of YEA with either method, with stable F statistics around 200, M7's below 0.2, and Q's below 0.5; these are highly seasonal series. From Table 2, the aggregate YEA has substantially better statistics with DA. For seasonal factors, the 85^{th} percentile is 3.5 vs. 5.3, and the maximum is 7.0 vs. 7.8; similarly, the statistics for month-to-month change favor DA. From Table 3,

Table 2. Sliding Spans Statistics

	Factors or adjusted series				Month-to-month change			
	85^{th}		max		60^{th}		max	
	DA	IA	DA	IA	DA	IA	DA	IA
YUN	1.6	1.2	3.9	4.2	1.2	1.4	5.0	5.6
AUN	1.2	0.8	2.8	1.3	0.7	0.7	3.2	1.7
FUN	0.9	0.9	2.9	1.5	0.8	1.0	3.7	2.0
MUN	1.0	1.1	1.8	1.9	0.8	0.8	2.1	2.1
YEA	3.5	5.3	7.0	7.8	2.5	4.5	7.8	12.1
AEA	1.0	0.6	2.1	1.4	0.7	0.7	1.9	1.8
FEA	1.4	1.7	3.6	4.2	1.2	1.1	3.8	4.4
MEA	1.1	0.8	2.4	1.4	0.6	0.6	2.0	1.6
YEN	0.7	0.7	1.3	1.6	0.5	0.6	1.5	1.5
AEN	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.2
FEN	0.1	0.1	0.3	0.3	0.1	0.1	0.3	0.3
MEN	0.1	0.2	0.3	0.3	0.1	0.1	0.2	0.3
UR	1.2	0.7	2.7	1.1	0.8	0.6	3.6	1.5

Table 3. Breakdown of Large MPD's

Series		2-3%	3-4%	4-5%	>5%
YEA	DA	18	11	4	5
	IA	14	18	7	19
YUN	DA	6	4	0	0
	IA	5	1	3	0
FEA	DA	6	4	0	0
	IA	4	1	1	0
FUN	DA	6	0	0	0
	IA	0	0	0	0

with IA, 19 months (23%) even exceed a "liberal" standard of 5% for differences in seasonal factors, compared to 5 with DA. Further support for DA comes from principle (3), the similarity of average seasonal patterns for the components YFEA and YMEA in Figure 1. Somewhat surprisingly, revision statistics are lower for IA, with average absolute revision 1.2% for 1996 vs. 2.1% for DA.

The values in Tables 2 and 3 also favor DA for YUN, teenage unemployment, but not by a very great margin. Its components also have a fairly similar average pattern, and the revision statistics favor DA. Dagum also suggested combining AMEA and AFEA. However, the sliding span and revision statistics are smaller with IA.

So far, aggregates by age have been analyzed. An aggregation scheme must take into account aggregates by gender, which also receive attention from users. Table 2 shows that IA performs as well or better for all except FEA. Even in this case, DA has more MPD's greater than 2, although IA has the largest MPD. A similar seasonal pattern for components AFEA and YFEA supports DA.

Looking at aggregates by both age and gender brings out the difficulty of reducing components for seasonal adjustment. Suppose we use YEA in place of its two components. FEA can either be adjusted directly or indirectly from YFEA and AFEA, but in either case total agricultural employment based on gender series may be inconsistent with the sum of AMEA, AFEA, and YEA. Principle (3) is not compelling when there is independent interest in the components or aggregates formed from them. Some pairs of components have different patterns. For example, AMUN has a single January peak, while AFUN, has peaks in both January and August. In summary, for most aggregates, IA provides sliding span (and revision) statistics which are as good or better than DA. Given that both age and gender aggregates are important, there appears to be no good case for combining any of the current set of 12 components of the unemployment rate.

As the foregoing analysis indicates, not all 12 of the current components are equally strong series. Part of the reason for both this study and Dagum's is the declining size of the agricultural sector, and the fact that CPS sample sizes are small for both agriculture and teenagers. One reason for the split between farm and nonfarm is that BLS's establishment employment survey covers nonfarm only. This, in itself, is probably not sufficient justification for using this employment breakdown. For the demographic variable age, alternatives could be considered. Dagum studies the age split 16-24 and 25+. As the percentage of individuals attending college increases, with a corresponding delay in stable, year-round employment for many, the 20-24 group may be more like the 16-19's than the 25+ group. A series with this alternative

Fig. 1. Average Seasonal Factors

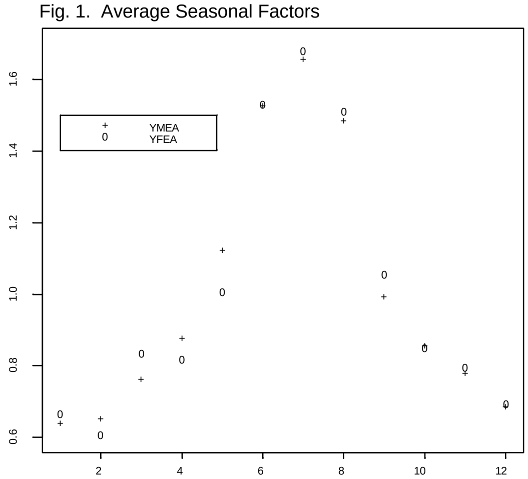
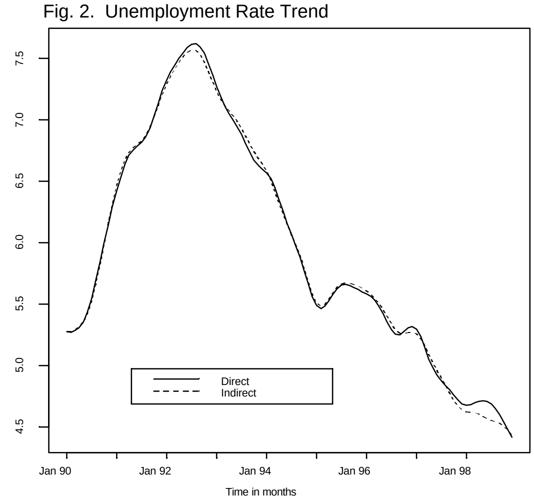


Fig. 2. Unemployment Rate Trend



"youth" classification should have lower variability than the current one. Finally, some other variable, such as race or ethnicity, may be important enough to merit consideration for contributing to the formation of the unemployment rate. The suggestions in this paragraph are largely the province of the economists and data users, and are beyond the scope of this paper. Given that some of the current components are rather weak series, however, it seems fitting to point out that alternatives are possible.

What about the unemployment rate itself? At that highly aggregated level, the statistics are quite good by both methods. IA does have smaller sliding span statistics in Table 2, and the graph of trends in Figure 2 does seem slightly smoother with IA. Thus, in addition to providing consistency, IA appears to give a slightly better seasonal adjustment for our most important statistic, the civilian unemployment rate.

3. Comparisons using X-11 and model-based seasonal adjustment

For this paper, TRAMO/SEATS (T/S) and X-12-ARIMA (X12), respectively, are used for model-based and X-11 seasonal adjustments. Comparing underlying methods has not been achieved by us, since X12 has more features and diagnostics, and is better known to us. What is done is to carry out an automatic seasonal adjustment with T/S for a series, and, starting from the T/S model, work with X12 to try to improve the ARIMA modeling part. This "user-tinkering" includes X12's transformation and outlier detection, examination of graphs, and comparison of results for closely related series. Thus, the aim of this section is not to gauge superiority of method or software. Instead, we address briefly the following questions:

(1) Are model estimation and outlier detection similar with T/S and X12?

(2) Does user-tinkering (with X12) improve on automatic adjustment (with T/S)?

To emphasize what is being compared, we use the notation T/A for automatic TRAMO/SEATS and X/U for X12 with user-tinkering. Table 4 summarizes modeling results for the 12 components and a few aggregates. T/S has a full-fledged model selection procedure, to be incorporated into X12, which is why we start with the T/A model. Both packages were asked to choose between a log transformation and no transformation, and to identify outliers. From the table, there is agreement on the choice of transformation, except for two series, YMUN and AFEA, where X12 chooses no transformation. T/A is geared very much to the airline model, choosing it for all but two series, YMUN and YEA. In these two cases, we accept the T/A model. For four series, where T/A chooses

Table 4. T/A and X/U Modeling Results

Series	Trans-formation		Model		q_q		q_{12}		No. of outliers		Box-Ljung p value	
	T/A	X/U	T/A	X/U	T/A	X/U	T/A	X/U	T/A	X/U	T/A	X/U
YFUN	L	L	011 011	011 000	.78	.72	.99	-	0	0	.21	.16
YMUN	L	N	210 011	210 011	-	-	.77	.62	3	5	.94	.15
AFUN	N	N	011 011	011 011	.41	.23	.99	.77	0	3	.43	.38
AMUN	L	L	011 011	013 011	.31	-.28	.72	.75	0	2	.02	.34
YFEA	L	L	011 011	100 011	.49	-	.93	.82	0	0	.14	.20
YMEA	L	L	011 011	300 000	.41	-	.94	-	0	0	.09	.20
AFEA	L	N	011 011	011 011	.31	.22	.59	.55	1	3	.96	.50
AMEA	L	L	011 011	011 000	.39	.31	.92	-	0	2	.24	.34
YFEN	N	N	011 011	011 011	.40	.32	.91	.79	0	0	.69	.57
YMEN	N	N	011 011	011 011	.31	.17	.82	.67	0	1	.28	.48
AFEN	L	L	011 011	011 011	.37	.37	.53	.44	0	3	.05	.18
AMEN	L	L	011 011	110 011	.10	-	.60	.57	0	1	.08	.15
YUN	L	L	011 011	011 000	.58	.52	.97	-	0	5	.19	.35
AUN	L	L	011 011	013 011	.16	-.25	.45	.42	0	4	.68	.86
YEA	L	L	100 011	100 011	-	-	.99	.97	0	0	.40	.23
AEA	N	N	011 011	011 011	.35	.35	.77	.77	0	0	.55	.37

an airline model, we use a deterministic seasonal, built in for X12, but not T/S. In all four cases, the seasonal MA parameter exceeds 0.90 with T/A, and is even closer to 1 with X12. However, after comparing results, keeping seasonal parameters very near one may be advantageous. For another series, YEA, we accept the airline model in X/U, even with $q_{12} = .97$. For a series such as AMUN, the rather low p-value for Q(24) from the airline model has caused us to seek a better-fitting model.

T/A selects outliers for only two series. The comparatively large number of outliers with X/U is mostly due to examination of graphs and tables, i.e., user-tinkering. Sometimes, this leads to sets of outliers which achieve the rather stringent standards of both packages; often, one or more outliers is selected with lower t-statistics. A possible advantage comes when including an outlier allows a more parsimonious model to have an acceptable fit. With AMEN, one outlier with X/U improves a bit on the marginal fit with T/A, with an equally parsimonious model. Probably, little harm is done with marginal outliers, at least when they are additive outliers; such values are also likely to be replaced by X-11 in its iterations.

The table exhibits differences, even with a common model, although these differences don't appear to be important. For YFEN, small differences appear in the estimation of both parameters and in the p-values. For AEA, the parameter estimates agree, but there is a difference in the p-values. Overall, more differences in modeling appear than we expected.

As mentioned above, in a rare departure from the airline model, T/A selects a (1 0 0) (0 1 1) model for YEA, teenage agricultural employment. Notice that in automatic mode, it selects no regular difference, a decision corroborated by examination of residual autocorrelation graphs. Average seasonal factors are quite close for the two methods. Figure 3 compares trends. Several countries are now publishing or doing research on trend estimates. Some of the research in this area involves the problem of false turning points. Examining trends may be a useful diagnostic, even if the trends are not published. Most people tend to prefer a smoother, simpler trend. In fact, both these methods seem to give pretty detailed trends, so smoother seems better. Overall, we like T/A's trend a little better. X/U's is smoother in the sense of being rounded. On the other hand, most of the time, its peaks and troughs are more extreme than T/A's. Here are a couple of hypotheses arising from this graph, which we would like to test further.

(1) T/S tends to give a more moderate trend, because, with the canonical decomposition, it tries to place as much as possible in the irregular.

(2) Locally X12's trend is smoother because its robustness feature point by point assigns certain variability to the irregular.

Figure 5 contains monthplots of the seasonal factors. While the averages are very close, the within-month patterns are quite different. In X/U's view, seasonal effects evolve. There is a fairly smooth curve for each monthly subseries. According to T/A, seasonal effects fluctuate. Maybe this latter view is better for some series.

At the seasonal adjustment conference in Bucharest, some comparisons of X12 & T/S used defaults for both. That made for a "fair" comparison, but in practice for important series such as these, we

Fig. 3. YEA Trend

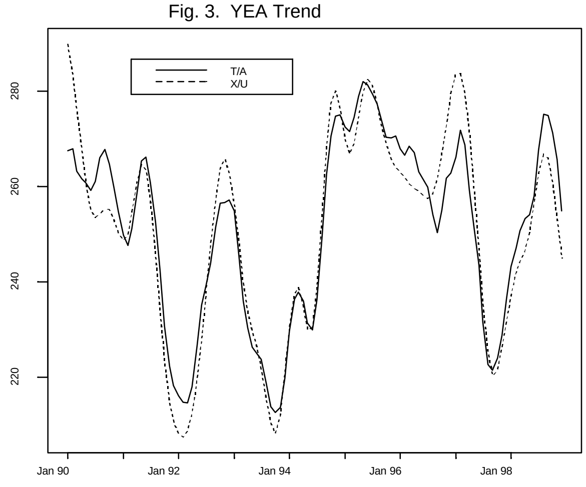


Fig. 4. AMUN Trend

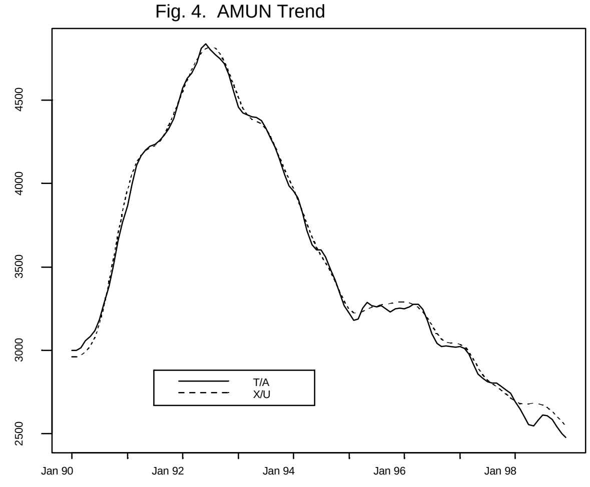
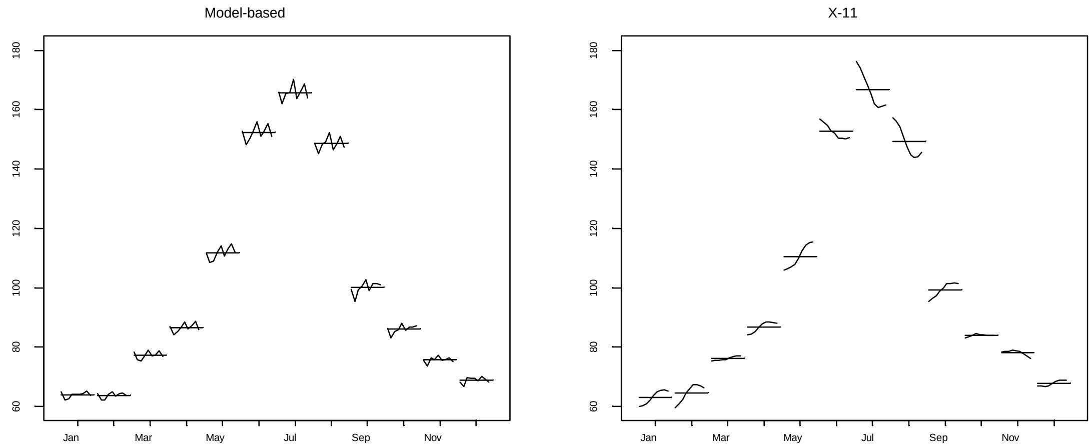


Fig. 5. Seasonal Factors for YEA



may want to use some of the tools that are available. Part of the reason for examining AMUN further is that the Box-Ljung statistic is not too good for T/A. Examining the residual autocorrelations leads fairly easily to the 013 011 model. Secondly, an additive outlier and a temporary change are included in the model, both with marginal t-statistics by the standards of both programs. Is there any difference? Figure 4 shows the trend graphs. They're quite close. However, to use Estela Dagum's term, there are more ripples in the T/A trend. The X/U trend appears preferable. The sliding span statistics (not shown) are a bit smaller with X/U. The monthplots (not shown) are similar; unlike the previous example, both generate evolving seasonal effects.

4. Summary

Indirect adjustment appears to perform as well or better than direct adjustment for most of the labor force aggregates studied, including the civilian unemployment rate. Direct adjustment seems better for teenage unemployment (YUN) and, especially, teenage agricultural employment (YEA), but its use would introduce inconsistency with gender aggregates, where indirect adjustment also performs well. This argues against any change in the method of adjustment with the current set of components. With small samples for the four components of teenage unemployment and teenage farm employment, it would be desirable to consider alternative components, such as a different age breakdown.

For two series, comparisons have been made between model-based and X-11 adjustments, with automatic TRAMO/SEATS and user-augmented X-12-ARIMA, respectively. Analysis of teenage agricultural employment shows that with model-based adjustment (1) seasonality fluctuates and (2) the trend has more moderate peaks and troughs, which may be desirable properties for some series. Analysis of adult male unemployment shows modest gains applying X-11 with user-tinkering.

Acknowledgments

Thanks to Catherine Hood for much help with TRAMO/SEATS and to Bob McIntire, Tom Nardone, and David Findley for information and useful conversations.

The views expressed are those of the authors and do not represent official positions of BLS.

References

Dagum, Estela Bee (1978), *A Comparison and Assessment of Seasonal Adjustment Methods for Employment and Unemployment Statistics*, Background Paper No. 5, U.S. National Commission on Employment and Unemployment Statistics, Washington, D.C.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen (1999), *X-12-ARIMA Reference Manual*, Version 0.2.3, Bureau of the Census, Washington, DC.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program," with discussion, *Journal of Business and Economic Statistics* 16, 127-177

Fischer, Bjorn and Christophe Planas (1998), "Large Scale Fitting of ARIMA Models and Stylised Facts of Economic Time Series," Eurostat Working Paper 9/1998/A/8, Eurostat

Gomez, Victor and Augustin Maravall (1997), *Programs TRAMO and SEATS: Instructions for the User*, Bank of Spain