ISSN Pending
Volume 1 number 1
May 1, 2016

Advanced Research on Economics and Management

Editor-in-chief: Bob Lew

1088 EMAIL PRESS

Advanced Research on Economics and Management

ISSN Pending

Volume 1, number 1, May 1, 2016

Impact Factor: Pending

Contents

Editorial Board

157 East ELM Street, Unit A, Greenwich, CT 06830-6614, USA

E-mail: arem@1088.email

Online first

Nonparametric Long Term Prediction of Stock Returns with Generated Bond Yields

Michael Scholz [1, *], Stefan Sperlich [2], Jens Perch Nielsen [3]

[1] Department of Economics, University of Graz, Universitätsstraße 15/F4, 8010 Graz, Austria

[2] Geneva School of Economics and Management, Université de Genéve, Bd du Pont d'Arve 40, 1211 Genévre 4, Switzerland

[3] Faculty of Actuarial Science and Insurance, Cass Business School, 106 Bunhill Row, London, EC1Y8TZ, UK

* Corresponding Author. E-mail: :michael.scholz@uni-graz.at

Abstract. Recent empirical approaches in forecasting equity returns or premiums found that dynamic interactions among the stock and bond are relevant for long term pension products. Automatic procedures to upgrade or downgrade risk exposure could potentially improve long term performance for such products. The risk and return of bonds is more easy to predict than the risk and return of stocks. This and the well known stock-bond correlation motivates the inclusion of the current bond yield in a model for the prediction of excess stock returns. Here, we take the actuarial long term view using yearly data, and focus on nonlinear relationships between a set of covariates. We employ fully nonparametric models and apply for estimation a local-linear kernel smoother. Since the current bond yield is not known, it is predicted in a prior step. The structure imposed this way in the final estimation process helps to circumvent the curse of dimensionality and reduces bias in the estimation of excess stock returns. Our validated stock prediction results show that predicted bond returns improve stock prediction significantly.

Keywords: Prediction; Stock returns; Bond yield; Cross validation; Generated regressors.

1. Introduction and motivation

For a long time predicting asset returns has been a main objective in the empirical finance literature. It started with predictive regressions of independent variables on stock market returns. Typically, valuation ratios are used that primarily characterise the stock, for example the dividend price ratio, the dividend yield, the earnings price ratio or the book-to-market ratio. Other variables related to the interest rate like treasury-bill rates and the long-term bond yield, or macroeconomic indicators like inflation and the consumption wealth ratio, are often incorporated to improve prediction. For a detailed overview we refer to the examples and discussion in Rapach et al. (2005) or Campbell and Thompson (2008).

In this paper, we take the actuarial long term view using yearly data, and focus on nonlinear relationships between a set of covariates. There are not many historical years in our records and data sparsity is of great importance in our approach. One could also use data of higher frequency as weekly or daily data, but one has to remember that the logistics of prediction is then very different. In our approach using yearly data bias might be of big importance while variance becomes less of an issue. In other words, the usual variance-bias trade-off depends on the horizon. An adequate model for monthly data might perform worse for yearly data and vice versa. The reason for the use of yearly data is our interest in actuarial models of long term savings and their possible econometric improvement (see e.g. Bikker et al., 2012, Guillen et al., 2013a and Guillen et al., 2013b, Owadally et al., 2013, Guillen et al., 2014, or Gerrard et al., 2014). Our favoured methodology of validating the fully nonparametric models that we employ for the long term yearly data also originates from the actuarial literature (see Nielsen and Sperlich, 2003).

The apparent predictability found by many authors was controversially discussed. As Lettau and Nieuwerburgh (2008) note, correct inference is problematic due to the high persistence of financial ratios, which have poor out-of sample forecasting power that moreover shows significant instability

over time. Therefore, the question of whether empirical models are really able to forecast the equity premium more accurately than the simple historical mean was intensively debated in the finance literature. Recently, Goyal and Welch (2008) fail to provide benefits of predictive variables compared to the historical mean. In contrast, Rapach et al. (2010) recommend a combination of individual forecasts. Their method includes the information provided from different variables and reduces this way the forecast volatility. Elliott et al. (2013) suggest a new method to combine linear forecasts based on subset regressions and show improved performance over the classical linear prediction methods. More recently, Scholz et al. (2015) propose a simple bootstrap test about the true functional form to evidence that the null of no predictability of returns can be rejected when using information such as earnings.

A direct comparison of stocks and bonds, mostly used by practitioners, makes the so-called FED model. It relates yields on stocks, as ratios of dividends or earnings to stock prices, to yields on bonds. Asness (2003) shows the empirical descriptive power of the model, but notes also that it fails in predicting stock returns. One of his criticisms is the comparison of real numbers to nominal ones. Actually, most studies discuss separately the predictability in stock and bond markets. However, Shiller and Beltratti (1992) analyse the relation between stock prices and changes in long-term bond yields. Fama and French (1993) find that stock returns have shared variation due to the stock-market factors, and they are linked to bond returns through shared variation in the bond-market. Engsted and Tanggaard (2001) pose the interesting question of whether expected returns on stocks and bonds are driven by the same information, and to what extent they move together. In their empirical setting, they find that excess stock and bond returns are positively correlated. Aslanidis and Christiansen (2014) adopt quantile regressions to scrutinise the realised stock-bond correlation and the link to the macroeconomy. Tsai and Wu (2015) analyse the bond and stock market responses to changes in dividends. Lee et al. (2013) find dynamic interactions among the stock, bond, and insurance markets. For additional literature on the relation between stock and bond returns (especially co-movements, joint distributions, or correlations), see, for example, Lim et al. (1998), Ilmanen (2003), Guidolin and Timmermann (2006), Connolly et al. (2010), Baele et al. (2010), or Bekaert et al. (2010).

One overall idea of the this paper is to exploit the interrelationship of present values of stock returns and bond returns. They are after all both discounted cash flows. Our underlying assumption implies that expected returns are associated with variables related to longer-term aspects of business conditions, as mentioned in Campbell (1987). Consequently, we include in a nonparametric prediction model of excess stock returns the bond yield of the same year. This way, the bond captures a most important part of the stock return, namely the part related to the change in long-term interest rate. Nonlinear forecasting methods are a growing area of empirical research, see for example Guidolin and Timmermann (2006), McMillan (2007), or Guidolin et al. (2009). Nielsen and Sperlich (2003) find a significant improvement in the prediction power of excess stock returns due to the use of nonlinear smoothing techniques. Based on their findings, we focus on nonlinear relationships between a set of covariates and the bond yield of the same year. We apply for estimation a local-linear kernel smoother which nests the linear model without bias. For the purpose of bandwidth selection and to measure the quality of prediction we use a cross-validation measure of performance. It is a generalised version of the validated R2R2 of Nielsen and Sperlich (2003) and allows for a direct comparison of the proposed model with the historical mean.

An obvious problem is that the current bond yield is unknown. Thus, we have to predict it in a first step. Here, we also employ fully nonparametric models and use a local-linear kernel smoother. This raises the question why it is necessary to use a two-step procedure. One could directly include the variables used for the bond prediction when forecasting stock returns. The problem is that such a model would suffer from the curse of dimensionality and complexity in several aspects: The dimension of the covariates, possible over-fitting, and the interpretability. In nonparametrics it is well known that the import of structure is an appropriate way to circumvent these problems.1 Furthermore, Park et al. (1997) showed that an appropriate transformation of the predictors can significantly improve nonparametric prediction. In our approach, we utilise the additional knowledge about the structure that is inherent in the economic process that generates the data. We find that the inclusion of

the generated variable shows notable improvement in the prediction of excess stock returns. Note that one does not achieve computational efficiency, but rather estimation efficiency from adding information. To our knowledge we are the first including nonparametrically generated regressors for nonparametric prediction of time series data. Therefore we also have to develop the theoretical justification for the use of constructed variables in nonparametric regression when the data are dependent.

For the empirical part we use annual Danish stock and bond market data (also used in Lund and Engsted, 1996, Engsted and Tanggaard, 2001, or Nielsen and Sperlich, 2003). We find that the inclusion of predicted bond yields greatly improves the prediction quality of stock returns in terms of the validated R2R2. With our best prediction model for one-year stock returns we not only beat the simple historical mean but we also observe a large increase in validated R2R2 from 5.9% to 28.3%. To underline our findings, we also include in our empirical analysis the prediction of the ratio of stock returns and dividend yields getting similar results.

The paper proceeds as follows. Section 2 describes the prediction framework and the measure of validation. The mathematical justification is introduced in Section 3. Section 4 presents our findings from an empirical and a small simulation study. Section 5 concludes. Finally, Appendix contains proofs of our theoretical results.

2. The prediction framework

In the financial and actuarial literature traditional approaches like the classic R2R2, the adjusted R2R2, goodness-of-fit or testing methods are mainly used to measure in-sample forecasting power. More recently, out-of-sample statistics and tests are discussed, see for example Inoue and Kilian (2004), Clark and West (2006), Goyal and Welch (2008), or Campbell and Thompson (2008). In our study, we use a generalised version of the validated R2R2 (View the MathML sourceRV2) of Nielsen and Sperlich (2003) based on leave-kk-out cross-validation. It measures how well a model predicts in the future compared to the historical mean. The classical R2R2 is often used, easy to calculate and has a straight forward interpretation. But it can hardly be used for prediction nor for comparison issues as it always prefers the most complex model. See also Valkanov (2003) or Dell'Aquila and Ronchetti (2006) for more relevant arguments for disregarding the classical R2R2 measure when selecting a model. For comparison often the adjusted R2R2 is applied, which penalises complexity via a degree of freedom adjustment. It is well known that this correction does not work in our case, see for example Sperlich et al. (1999).

The idea of the View the MathML sourceRV2 is to replace total variation and not explained variation by their leave-kk-out cross-validated analogs. Note that cross-validation (cv) is a quite common in the nonparametric time series context, see Györfi et al. (1990). More formally, consider the two models

The last term in (4) vanishes as we will see in Theorem 3.7 and the second term can be easily approximated. The gain in our two-step procedure comes now from the fact that the bond in the second term in (5) is quite predictable. We confirm this fact in the empirical part 4.2 (see Table 2). In the same vain, Lin et al. (2014) find that bond returns are more predictable then stock returns. Another idea would be the following: first, estimate gg with the available bond data ~~bt~~1bt−1, and second, evaluate View the MathML sourcegˆ at the constructed View the MathML sourcebˆt. Since, however, this procedure did not improve the stock forecasts, we skip it from further considerations.

One could directly use the variables in the vector ~~wt~~1wt−1 as regressors in model (2). But the model would suffer from complexity and dimensionality in several aspects: The dimension of the covariates as well as their interplay. In the nonparametric literature, typically two strategies are proposed to circumvent these problems-either semiparametric modelling or additivity, both to import structure. Nielsen and Sperlich (2003) showed that additive models fail to improve the prediction of stock returns due to a non-ignorable interaction between the predictors. We improve these results by providing additional structure which is inherited by the underlying data generating process. We think of the same years bond yield as an important factor which captures some of the relevant features for the expected stock returns. Then, the inclusion of bond yields when predicting stock returns

nonparametrically acts as a kind of complexity and dimension reduction due to the import of more structure.

To see if it is possible to further improve the predictive power in our setting, we will also analyse the model (2) with a different dependent variable. We consider the ratio between current stock returns and dividend yield, i.e. View the MathML sourceYt∗=Yt/dt (see Section 4).

3. Mathematical justification

We prove the consistency of a function estimate which makes use of constructed variables and derive its asymptotic properties. For the prediction in the time series context, we follow the steps from Ferraty et al. (2001) and combine them with Sperlich (2009).2 Let us consider a sample of real random variables {(Xi,Yi),i=1,⋯,n}{(Xi,Yi),i=1,⋯,n} which are not necessarily independent and want to estimate the unknown function m(x)=E(Y|X=x)m(x)=E(Y|X=x), x∈Rx∈R, that should always exist. Note that for time series {(Zi),i∈N}{(Zi),i∈N} a kk-step ahead forecast is included in a natural way setting Yi=Zi+kYi=Zi+k and Xi=ZiXi=Zi. We concentrate only on the case of an auto-regression function of order one. Since we face constructed realisations for XX, we assume a predictor3 with an additive bias and a

For mean square convergence, asymptotic normality and higher order polynomials, one could directly extend the work of Masry and Fan (1997) to the case of predicted regressors.

4. Empirical evidence and simulation studies

We interpret our method presented as a two stage regression approach. Based on the idea that the bond of the same year captures an important part of the stock return we search in the first step the optimal prediction model for the bond. Afterwards, as we have seen in Theorem 3.7, we can consistently predict stock returns using the predicted bond yields.

4.1. Data description

Consider the annual Danish stock and bond market data for the period 1923–1996 from Lund and Engsted (1996). In the Appendix of their work, a detailed description of the data can be found. We use a stock index based on a value weighted portfolio of individual stocks chosen to obtain maximum coverage of the market index of the Copenhagen Stock Exchange (CSE). Notice that the CSE was open during the second world war. When constructing the data, corrections were made for stock splits and new equity issues below market prices. Table 1 presents summary statistics of the available variables. In the following, we use the dividend price ratio, dd, the stock return, SS, the long-term interest rate, LL, the short-term interest rate, rr, and the bond yields, bb, as explanatory variables.

4.2. The prior step: a simple bond yield predictor

We speak of a simple predictor as in the literature quite complex models can be found for this problem. Our main target, however, are the stock returns where bond yield prediction is just an auxiliary step in order to reduce complexity and dimension. Therefore, the model and bandwidth selection for (3) has to be based on the objective of maximising the View the MathML sourceRV2 of the stock return problem (2). Recognising that the model that maximises the View the MathML sourceRV2 for bond prediction is not necessarily the one that maximises the View the MathML sourceRV2 for stock returns, it becomes clear that it is worth to consider nonparametric alternatives for (3), even if parametric models seem to do a very good job for bond yield prediction alone. This is the reason why we need Theorem 3.7; for parametric predictors View the MathML sourcex^ the consistency of (7) follows trivially.

If we just look at the bond yield prediction, then we get mostly positive View the MathML sourceRV2 for the models listed in Table 2. We observe that only in few cases a local linear predictor does a better job than a linear model as far as we look at the View the MathML sourceRV2 for bond yields. The interesting numbers, however, we will see only when looking at the View the MathML sourceRV2 for stock returns in Table 3, next section.

Fig. 1 shows the estimated functions View the MathML sourcepˆ for the bond yield prediction step with a single covariate wt−1wt−1 from the set {S,r,d,L}{S,r,d,L} using a linear model (triangles) and a fully nonparametric model (diamonds). For some of the models a clear nonlinear behaviour can be observed. Fig. 2 displays the estimation results of the combination of the variables r,br,b that gives the largest validated R2R2 value for 2-dim. models in Table 2, again for the linear model (triangles) and the fully nonparametric model (diamonds). Note that we set one variable at a certain level (25%, 50%, 75% quantile) and plot the relationship of bond yields with the remaining variable. For example, on the left-hand side of Fig. 2, we set the lagged bond yield at values of 2.0, 5.7, and 12.2. The linear model and the fully nonparametric model behave very similar (what is not surprising, since both have more or less the same validated R2R2 value). Only at the boundaries a clear difference of both models is visible. Note again that we are interested in stock return prediction and that the predicted bond used in the final step not necessarily has to be the best possible one.

4.3. Stock prediction

Now we examine the predictive relationship of excess stock returns YtYt and a set of financial variables vt−1vt−1 using different models. Results of this exercise are summarised in Table 3. First, for the sake of illustration, we develop our strategy step by step and start with the estimation of the model Yt=g(vt−1)+εtYt=g(vt−1)+εt with a fully nonparametric kernel based method as well as the parametric counterpart (not including the constructed bond yield View the MathML sourcebˆt). Part (a) of Table 3 reports the results and shows that all parametric models produce negative validated View the MathML sourceRV2 values. It means that with a linear regression approach we cannot better forecast one-year stock returns than the simple mean. A more sophisticated technique is needed. In fact, our so far best nonparametric model7 uses actual lagged bond yields, bt−1bt−1, and gives an View the MathML sourceRV2 of 5.9%. But even better results are possible when we include the generated bond yield View the MathML sourcebˆt in our analysis.

Second, we follow our procedure proposed in Section 2 and generate the current bond yield with model (3). Then we include this constructed variable as a regressor in the final step, the model for excess stock returns as stated in Eq. (2). Let us do this first without any further regressor vt −1vt−1. As discussed before, we have to choose the model and bandwidths along the largest View the MathML sourceRV2 value for predicting stock returns.8 How much the predictive power has increased by this method can be seen when comparing part (a) with part (b) of Table 3. The best model in (b) uses as single regressor lagged bond yields in the first step and only the predicted bond as covariate in the second step (View the MathML sourceRV2 of 10.6%). Even for the parametric counterpart our strategy helps to improve prediction power since we can observe positive View the MathML sourceRV2 for some models. As one can clearly see, the nonparametric version produces better results, recall our discussion in the previous section.

Third, we construct the current bond as before but accompany this regressor in model (2) by any combination of lagged variables from the predictor set {d,S,L,r,b}{d,S,L,r,b} as our vector vt −1vt−1. Then, the two largest View the MathML sourceRV2 were achieved by View the MathML sourcegˆ(bˆt,dt−1,St−1,Lt−1) where View the MathML sourcebˆt=pˆ(dt−1) (yielding View the MathML sourceRV2=30.3%) or View the MathML sourcebˆt=pˆ(dt,Lt−1) (yielding View the MathML sourceRV2=28.9%), respectively. Note that for an increasing set of regressor variables the corresponding multidimensional bandwidth grid on which we looked for the best predicting one had to be reduced for numerical reasons. Consequently, lower dimensional models have the tendency to be slightly favoured in our study. The full set of results for the 25 times 25 combinations of {d,S,L,r,b}{d,S,L,r,b} is not shown for the sake of presentation, but available on request. We include in part (c) of Table 3 only the ‘diagonal’ of those results since the predictive power is among the best of all possible models. In other words, in part (c) of Table 3 holds wt −1=vt−1wt−1=vt−1, exactly the same regressors used for the bond construction in step one accompany View the MathML sourcebˆt again in the second step. For completeness, also the results of the parametric counterpart are included in part (c) of Table 3. We see that our new prediction procedure improves the predictive power for stock returns. We find again convincing evidence that the two nonparametric steps are better than the parametric counterpart. For the best model in Table 3–we have

wt−1=vt−1=wt−1=vt−1=(dt−1,Lt−1)(dt−1,Lt−1)–we find a large increase in the View the MathML sourceRV2 value from 5.9% to 28.3%, an about factor five improvement compared to the best model without constructed bonds. This finding again indicates that the bond captures a quite important part of the stock return which is related to the change in long-term interest rate.

Fig. 3 shows the estimated functions View the MathML sourcegˆ for the excess stock return predictions based on a single covariate from the set {S,r,d,b}{S,r,d,b} using a fully nonparametric model (diamonds), a fully nonparametric model with the constructed bond as single regressor in the second step (crosses), and a fully nonparametric model based on the predicted bond together with the regressor of the first step (pluses). Again, for some of the models a clear nonlinear pattern can be observed. Fig. 4 displays the estimation results of the combination of the variables d,Ld,L that gives the largest validated R2R2 value in Table 3, again for three different models used in Fig. 3. Note that we set one variable at a certain level (25%, 50%, 75% quantile) and plot the relationship of excess stock returns with the remaining variable. For example, on the left-hand side of Fig. 4, we set the lagged long-term interest rate at values of 5.1, 6.4, and 10.6. It seems that the model which uses only the predicted bond as a covariate is too inflexible in its functional form and needs the additional information which is still inherent in the covariates of the first step when it comes to stock return prediction in step 2. This underlines the findings of Table 3 where the largest validated R2R2 values can be found in part (c) for the nonparametric estimators of the models based on the predicted bond together with the same covariates of the first step.

In order to get a better view of the potential of our proposed method, we analyse the out-of-sample mean-squared error (oos-mse) for a one-step ahead prediction with an expanding estimation sample for the new method in comparison to the oos-mse of the corresponding fully nonparametric and linear models as well as the historical mean. For an illustration we use dividend by price and long-term interest rate since we observed for this combination of covariates the largest View the MathML sourceRV2 in Table 3. Fig. 5 shows the predicted annual excess stock returns of the different models in comparison to the realised annual excess stock returns of the CSE. We observe the smallest oos-mse for the new method (0.044), followed by the historical mean (0.049), the linear model (0.058), and the fully nonparametric model (0.059).

The first line refers again to the parametric version of model (2) and the second line to the fully nonparametric method, both without constructed bonds. Almost all of the parametric models have negative View the MathML sourceRV2 values and also only a small number of nonparametric models beat the simple mean. In contrast, when we include the constructed bond in the nonparametric prediction, a large increase of the validated View the MathML sourceRV2 can be observed. For example, the model which uses long- and short-term interest rate, and lagged bond yields for both the bond generation and following stock prediction, has a View the MathML sourceRV2 value (43.5%) that is over three and a half times larger than the value of the best model without constructed bonds (12.3%).

4.4. Simulation studies

A simulation study gives us the possibility to highlight the potential of our method. We first show the effects of a dimension reduction and afterwards of a pronounced curvature.

Let us consider a four dimensional function that is separable into two terms: View the MathML sourcem(x1,…,x4)=m˜(s1,s2) with s 1=s1(x1,x2)s1=s1(x1,x2) and s 2=s2(x3,x4)s2=s2(x3,x4). We simulated data from the following models : S1=x1+x1x2+εσs1S1=x1+x1x2+εσs1, S2=exp(x3+x4)+εσs2S2=exp(x3+x4)+εσs2, and View the MathML sourceY=m(x1,…,x4)+εσm=m˜(s1,s2)+εσm=s1+s2+εσm. For each explanatory the support is [0,1][0,1]. An autoregressive design with ϕ=0.75,0.2,0.02ϕ=0.75,0.2,0.02 for x1,…,x3x1,…,x3 was used; also a normal for x4x4. Different parameter values σσ for the zero mean normal error distributions were investigated as well as different sample sizes nn. The kernel used was the Gaussian. For computational reasons the bandwidths are chosen separately in each step of the simulation.9 In step one we predict s1s1 and s2s2, used in step 2 to estimate function View the MathML sourcem˜.

Lines three and four of Table5 present the results for the two-step approach for View the MathML source$\tilde{m}$ and the fully nonparametric method estimating mm in terms of View the MathML sourceRV2 values, averaged over 500 runs. The proposed two-step procedure succeeds in improving on the fully nonparametric estimator in all cases by far. The effect of the dimension reduction is of course more pronounced for the smaller sample size and results in an almost factor 2 improvement.

For the second part we consider the function composition View the MathML source$m(x)=\tilde{m}\circ s(x)$, where the inner function ss has a pronounced curvature. We simulated data from the following models: $S=\sin(4\pi(x-1/8))+\cos(4/3\cdot\pi(x-1/2))+1.6+\varepsilon\sigma sS=\sin(4\pi(x-1/8))+\cos(4/3\cdot\pi(x-1/2))+1.6+\varepsilon\sigma s$ and View the MathML source$Y=m(x)+\varepsilon\sigma m=\sin\circ s(x)+\varepsilon\sigma m$, i.e. View the MathML source$\tilde{m}(x)=\sin(x)$. Note that ss is one of the example functions used in Park et al. (1997). A uniform design was used with the support [0,1][0,1]. Different parameter values σσ for the zero mean normal error distributions were investigated for a sample size of n=50n=50. The kernel used was the Gaussian and the bandwidths are chosen separately in each step for the two-step part. Again we are aware of the suboptimality, i.e. we could even do better with respect to the View the MathML sourceRV2 but at the cost of computing time. Table 6 reports the results. We find that already in this simple example the proposed two-step approach can help to obtain clearly better results, i.e. much larger View the MathML sourceRV2-values in all cases. Fig. 6 shows the used inner function (left) and estimates of mm and View the MathML source$\tilde{m}$ (right). We see that our method can better estimate problematic regions, in particular by bias reduction.

5. Concluding remarks and outlook

Motivated by economic theory and statistical arguments, we include the same years bond yield in the fully nonparametric prediction approach for excess stock returns. Since the current bond yield is unknown, we propose to construct it in a prior step using again nonparametric techniques. The bandwidths should be chosen in such a way that they maximise the View the MathML sourceRV2 of the final step. The empirical study demonstrates that this two-step approach can improve the stock return prediction enormously. We moreover prove the consistency of our method and derive the asymptotic behaviour of our final predictor. We illustrate the improvement due to our method using annual Danish stock and bond market data which were studied in detail in former articles by different authors. Our results confirm our motivation of including the same years bond yield, namely that it captures the most important part of the stock return, that one related to the change in long-term interest rate. This actually holds not only for stock returns but also for transformed variables, as for example returns divided by dividend yields.

The statistically insights are the following. It is clear that we face a regression model that exhibits high complexity and dimensionality. An obvious remedy would be the imposing of structure. Since it has been shown that additive separability is inappropriate because of unknown interactions, we make use of financial theory to exploit the inherit structure of stock returns. Alternatively, one could interpret the first stage as an optimal nonparametric transformation that maps, for example, the long-term interest rate to the current bond yield, View the MathML source$Lt\ \mapsto\hat{b}t$. The subsequent nonparametric smoother of the transformed variable is than characterised by less bias. Here, we present a practical example in the spirit of the somewhat theoretical method proposed by Park et al. (1997) which improves nonparametric regression with transformation techniques. Although we extend their method in several aspects, their paper provides some statistical intuition for the success of our approach. Our simulations additionally underpin the key idea of complexity and dimension reduction.

References

1. Aslanidis and Christiansen, 2014. N. Aslanidis, Ch. Christiansen. Quantiles of the realized stock–bond correlation and links to the macroeconomy. J. Empir. Finance, 28 (2014), pp. 321–331.
2. Asness, 2003. C. Asness, Fight the fed model: The relationship between future returns and stock and bond market yields. J. Portf. Manag., 30 (1) (2003), pp. 11–24.

Jean-Denis Garon. Adv. Res. Econ. Manage., 2016, 1(1): 8-11.

The Commitment Value of Funding Pensions

Jean-Denis Garon [1, 2, 3, *]

[1] ESG-UQAM, Canada

[2] CIRPÉE, Canada

[3] CESIfo, Germany

* Corresponding Author: Douglas Scott Reintgen, MD, Department of Surgery, University of South Florida Morsani School of Medicine, MDC 52, 12901 Bruce B., Downs Boulevard, Tampa, FL 33612.
E-mail contact: dreintge@health.usf.edu

Abstract. This paper studies how funding public pensions can improve policy outcomes when short-sighted governments cannot commit. We focus on sustainable plans, where optimal nonlinear pensions are not reneged on by sequential governments. Funding pensions is a commitment mechanism. It implies lower contributions than does the second best policy, which reduces temptation to over-redistribute later and to misuse revealed private information. Funding may be preferable even if the population growth rate is higher than the rate of return on assets. Second best optimal policies are also more likely to be renegotiation proof under fully funded pensions.

Keywords: Pensions; Commitment; Redistribution; Funding.

1. Introduction

Publicly managed pension plans are subject to political risks (Diamond, 1994 and Diamond, 1996). Even benevolent governments may be tempted to engage in excess redistribution among retirees using pension wealth. Because of this, some have argued that funding and privatizing public pensions could reduce political risks.

Recent literature in dynamic optimal taxation, among which Farhi et al. (2012), has shown that commitment is especially relevant in dynamic non-linear optimal tax problems, in which the fiscal schedule must induce individuals to reveal private information about themselves. If the policy maker can improperly use revealed information and renege on its promises, the optimal policy may be significantly altered and capital should be taxed progressively. Doing so reduces income inequality in the optimum. Sequentially, governments thus have fewer incentives too misuse households' private information to over-redistribute. Farhi et al. (2012) study sustainable equilibria à la Chari and Kehoe (1990) that are perfect Bayesian and that can be sustained by a trigger-type reaction by the households following a governmental deviation.

We extend their analysis to show how the institutional structure of public pensions, whether fully funded or unfunded, may help or harm policy outcomes when commitment is assumed away. We use a simple, overlapping generations model, with an infinite repeated game between successive governments and generations. An initial social planner who sets contribution levels and the redistributive characteristics of the public pension plan must ensure that successive short-sighted governments do not have an incentive to renege later on.

Our results formalize the idea that funding pensions may be used as a commitment mechanism. When it is, the optimal response to a lack of commitment is to reduce aggregate pension contributions in order to reduce next period's temptation. With unfunded plans, immediate temptation to over redistribute involves higher contributions than in the second best plan, and significantly less inequality. We use numerical examples to show that optimal second best policies are more likely to be sustainable under funded pensions. Due to their pre-commitment value, funded pensions may be preferable to pay-as-you-go schemes even when the rate of return on financial assets is smaller than the growth rate of the population.

2. Model

Consider an overlapping generations version of Stiglitz (1982) where individuals live for two periods of equal duration. In the first half of their lives they supply labor, consume, are taxed and contribute to a public pension fund. In the second half they are retired and live off public pension benefits. The timing of retirement is exogenous and population grows at a fixed rate $\eta>0\eta>0$. Thus, at each period $t=0,1,\ldots t=0,1,\ldots$ one generation of workers cohabits with one generation of retirees. The constant ratio of workers to retirees is therefore $1+\eta 1+\eta$. There is a constant proportion $n_i n_i$ of type-i agents, where types are denoted by $i=1,2i=1,2$. There is an underlying linear production technology according to which a type-i worker who supplies View the MathML source$\ell_t i$ units of labor faces a hourly market wage rate $w_i w_i$ with $w1<w2w1<w2$. Gross incomes are defined as View the MathML source$y_t i \equiv w_i \ell_t i$. All individuals have identical, time separable utility functions:\

By (3a) aggregate consumption of workers equals aggregate gross income minus pension contributions. By (3b) aggregate consumption of retirees depends on $\alpha \in \{0,1\} \alpha \in \{0,1\}$, which captures whether public pensions are unfunded fully funded. Following the taxonomy of Lindbeck and Persson (2003), in an unfunded plan ($\alpha=0\alpha=0$) aggregate benefits are financed by a specific tax on the generation currently working. A fully funded plan ($\alpha=1\alpha=1$) has them financed by the returns on previously accumulated pension assets. For simplicity, assume that these savings yield the fixed rate of return r, as one would find in a small open economy.1 Note that a fully funded pension plan need not be actuarially fair at the individual level because individual benefits are not necessarily proportional to one's own contributions (Feldstein and Liebman, 2002 and Lindbeck and Persson, 2003). Note finally that $\alpha\alpha$ is taken as an institutional feature. Since it is fixed, it is highly costly to reform on short notice. It captures the stylized fact that pension contribution rates are more frequently adjusted than the fundamental structure of public pension plans, which requires in-depth reform, more time and more policy debates to implement than simply changing contribution rates.

2.1. Full commitment benchmark

Choosing an optimal allocation is equivalent to designing a nonlinear tax system across workers and retirees. Suppose that at $t=0t=0$ the social planner can once and for all promise future allocations that satisfy the feasibility constraints. He maximizes (2) by choosing View the MathML source$\phi_t, \forall t$ subject to and .2 Unsurprisingly, concave utility of consumption (or aversion to inequality) prescribes View the MathML source$c_t 1=c_t 2, d_t 1=d_t 2$, and View the MathML source$y_t 1<y_t 2 \forall t$. All individuals have identical consumptions, but type-2s are invited to work more (Mirrlees, 1971 and Stiglitz, 1982).

As is well known since Mirrlees (1971), such an allocation is not incentive compatible. If only gross incomes View the MathML source$y_t i$ can be observed instead of types, type-2 workers will mimic type-1s. Second best optimality is therefore restricted to incentive compatible allocations that satisfy

2.2. Sequential governments

Suppose now that the social planner initially promises allocations View the MathML source$\phi_t, \forall t$. Each allocation must be incentive compatible and feasible. Lagrange multipliers View the MathML sourceθ_t, μ_t, and $\lambda t \lambda t$ are assigned to Eqs. (4), and . However, the social planner does not have the final say. Sequential governments can later re-optimize and change allocations insofar as they are feasible. We model them in the spirit of Farhi et al. (2012), where three motives induce sequential governments to renege. First, they already know retirees' types and may seek to set View the MathML source$d_t 1=d_t 2$. Second, they may weigh generations differently than does the initial social planner. Third, accumulated assets are perceived as an inelastic tax base that can be redistributed at no immediate efficiency cost. The objective function of a time tt government is

where $\pi\pi$ is the weight put on current retirees, whose types are known.

Let us focus on allocations that can be promised by the planner at $t=0t=0$ and which sequential governments will not renege on. Oftentimes, such policies have been characterized by taking the limit of the backward induction solution to a dynamic game.3 Here, young workers know that their private information will be used to equalize consumption across types next period. Therefore, they reveal it only if the promised allocations maximize (6) subject to View the MathML source$d_t 1=d_t 2=d_t$, to the feasibility constraint and to the IC constraint.

To separate types, allocations must therefore allow for more inequality across workers . But what interests us is the role of accumulated assets btbt on the outcome of this game . Denoting View the MathML sourceσ¯t the allocation selected by governments (and promised by the social planner) and View the MathML sourceW¯t a government's value function, we find that4

We first study a scenario where r=η=1r=η=1. If ππ is low enough so the credibility constraint does not bind at the second best allocation, the funded and unfunded regimes yield a social welfare of approximately 9.6321 in the steady state. Inter generational discounting implies that retirees' consumption is marginally higher under the unfunded regime.

When pensions are unfunded, increases in ππ quickly translate into binding temptation. Since no commitment mechanism is available, sequential governments react by increasing retirees' consumption. Otherwise, reneging would take place later on. Social welfare decreases to eventually attain negative values.

With funding, a broader range of second best allocations can be sustained without commitment. Only when ππ goes from 0.60 to 0.65 does the credibility constraint bind. When it does, one can readily see how pre-commitment kicks in. Instead of increasing contributions (as under unfunded pensions), the social planner reduces them to diminish sequential governments' amount of cash-on-hand that is available for redistribution.

The second numerical example is found in the four rightmost columns of Table 1. We have set η>rη>r and adjusted ρρ so that δδ remains unchanged. In our setup, η>rη>r makes the unfunded regime strictly dominant in any full commitment scenario since we abstract from risk issues (Dutta et al., 2000). However, as ππ increases the commitment value of funding pensions makes it preferable not to resort to pay-as-you-go schemes.

3. Conclusion

We used a simple model to explain why funding pensions may help governments to commit. Our stylized assumptions helped us formalize why and when funding may act as a commitment device. For simplicity we assumed that the funding structure of the pension plan is difficult to change in the short run, whereas the contribution rates can be more easily adjusted. Also, fully funding the pension plan (in a steady state) implies, in the model, that policy instruments that redistribute across generations are shut down. In this environment, the normative case for fully funded pensions becomes stronger when commitment is also assumed away.5

Of course, more still has to be done on this topic. Some researchers, such as Blake (2000) and Barr (2002), contend that funded pensions are at best an imperfect commitment device to isolate pension capital from political risks.6 While governments can (and do) break their PAYG promises, they can equally reduce the real return to pension funds, by requiring fund managers to hold government financial assets with a lower yield than they could earn elsewhere, or by withdrawing or reducing any tax privileges. The Argentinean case also convincingly demonstrates that simply ending pay-as-you-go schemes and transferring pension management to the private sector does not mechanically alleviate political risks (Kay, 2009). A new set of political risks can then emerge since funded assets can be perceived as an inelastic tax base by predatory and short-sighted governments, with excess redistribution and time-inconsistent policy-making as consequences.

Acknowledgments

Lymphatic mapping and SLNB will continue to play an important role in the treatment of women with breast cancer. Although some controversy exists in determining the effect of nodal staging on the treatment and prognosis of women, the knowledge gained from the technique continues to be used to guide therapy and determine the prognosis. The SLN procedure will evolve by eliminating the need for radioactivity in the operating room, and the technique will become more accurate and used in expanded indications by incorporating preoperative imaging and intraoperative procedures.

References

1. Cabanas RM. An approach for the treatment of penile carcinoma. Cancer 1977; 39:456-66.

2. Morton DL, Wen DR, Wong JH, et al. Technical details of intraoperative lymphatic mapping for early stage melanoma. Arch Surg 1992; 127:392-9.
3. Krag DN, Weaver DL, Alex JC, et al. Surgical resection and radiolocalization of the sentinel lymph node in breast cancer using a gamma probe. Surg Oncol 1993; 2:335-9.
4. Norman J, Cruse W, Ruas E, Beatty E, Reintgen DS. The expanding role of lymphoscintigraphy in malignant melanoma. Am Surg 1989; 55: 689-98.
5. Reintgen DS, Cruse CW, Wells K, et al. The orderly progression of melanoma nodal metastases. Ann Surg 1994; 220:759-67.
6. Albertini JJ, Lyman GH, Cox C, Reintgen DS. Lymphatic mapping and sentinel lymph node biopsy in the patient with breast cancer. JAMA 1996; 276:1818-22.
7. Boolbol SK, Fey JV, Borgen PI, et al. Intradermal isotope injection: a highly accurate method of lymphatic mapping in breast carcinoma. Ann Surg Oncol 2001; 8: 20-4.
8. Cody HS, Frey J, Akhurst T, et al. Complementarity of blue dye and isotope in sentinel node localization in breast cancer: univariate and multivariate analysis of 966 procedures. Ann Surg Oncol 2001; 8:13-9.
9. Hill AD, Tran KN, Akhurst T, et al. Lessons learned from 500 cases of lymphatic mapping for breast cancer. Ann Surg 1999; 299:528-35.
10. Yalom M. A history of the Breast. New York: Alfred A. Knopf; 1997.
11. Halstead W. The results of operation for cure of cancer of the breast performed at John Hopkins Hospital. Johns Hopkins Hosp Bull 1894; 4:497.
12. Devitt JE. The influence of conservation and radical surgery on survival of patients with breast cancer. Can Med Assoc J 1962; 87:906-10.
13. Fisher B, Montague E, Redmond C, et al. Comparison of radical mastectomy with alternative treatments for primary breast cancer: a first report of results from a prospective randomized clinical study. Cancer 1977; 39:2827-39.
14. Euhus DM. Are axillary lymph nodes still relevant in breast cancer? Ann Surg Oncol 2014; 21:4051-3.
15. Reintgen M, Murray L, Akman K, et al. Evidence for a better nodal staging system for melanoma: the clinical relevance of metastatic disease confined to the sentinel lymph nodes. Ann Surg Oncol 2014; 20:668-74.
16. Dessureault S, Soong SJ, Ross MI, et al. AJCC Melanoma Staging Committee. Improved staging of node-negative patients with intermediate to thick melanomas (>1 mm) with the use of lymphatic mapping and sentinel lymph node biopsy. Ann Surg Oncol 2001; 8:766-70.
17. Protic M, Stojadinovic A, Nissan A, et al. Prognostic effect of ultra-staging nodenegative colon cancer without adjuvant chemotherapy: a prospective National Cancer Institute-sponsored clinical trials. Ann Surg Oncol 2015; 22:3296-301.
18. Boughey JC, Ballman KV, Hunt K, et al. Axillary ultrasound after neoadjuvant chemotherapy and its impact on sentinel lymph node surgery: results from the ACOSOG Z1071 trial (Alliance). J Clin Oncol 2015; 33:3386-93.
19. Mittendorf EA, Ballman KV, McCall LM, et al. Evaluation of the stage IB designation of the American Joint Committee on Cancer staging system in breast cancer. J Clin Oncol 2015; 33:1119-27.
20. Querzoli P, Pedriali M, Rinaldi R, et al. Axillary lymph node nanometastases are prognostic factors for disease-free survival and metastatic relapse in breast cancer. Clin Cancer Res 2006; 12:6696-701.
21. de Boer M, van Deurzen CH, van Dijck JA, et al. Micrometastases or isolated tumor cells and the outcome of breast cancer. N Engl J Med 2009; 361: 653-63.
22. Weaver DL, Ashikaga T, Krag DN, et al. Effect of occult metastases on survival in node-negative breast cancer. N Engl J Med 2011; 364:412-21.
23. Shigematsu H, Taguchi K, Koui H, Ohno S. Clinical significance of extracapsular invasion at sentinel lymph nodes in breast cancer patients with sentinel lymph node involvement. Ann Surg Oncol 2015; 22:2365-71.
24. Swaminathan S, Reintgen M, Kerivan L, Smith J, Reintgen D. Extracapsular extension in the sentinel lymph node: recommendations for therapy. Clin Breast Cancer (in press).
25. Krag D, Weaver D, Ashikaga T, et al. The sentinel lymph node in breast cancer—a multicenter validation study. N Engl J Med 1998; 339:941-6.
26. Giuliano AE, Kirgan DM, Guenther JM, et al. Lymphatic mapping and sentinel lymphadenectomy for breast cancer. Ann Surg 1994; 220:391-8.
27. O'Reilly EA, Prichard RS, Azawi DA, et al. The value of isosulfan blue dye in addition to isotope scanning in the identification of the sentinel lymph node in breast cancer patients with a positive lymphoscintigraphy: a randomized controlled trial (I; 127:SRCTN98849733). Ann Surg 2015; 262:243-8.
28. Kim T, Giuliano AE, Lyman GH. Lymphatic mapping and sentinel lymph node biopsy in early-stage breast carcinoma: a meta-analysis. Cancer 2006; 106:4-16.
29. Bezu C, Coutant C, Salengro A, et al. Anaphylactic response to blue dye during sentinel lymph node biopsy. Surg Oncol 2011; 20:55-9.
30. Intraoperative fluorescence imaging for sentinel lymph node detection: prospective clinical trial to compare the usefulness of indocyanine green vs. technetium Tc99m for identification of sentinel lymph nodes. JAMA Surg 2015; 150:617-23.
31. de Boer M, van Dijick JA, Bult P, et al. Breast cancer prognosis and occult lymph node metastases, isolated tumor cells and micrometastases. J Natl Cancer Inst 2010; 102:410-25.
32. Agrawal A, Civantos FJ, Brumund KT, et al. [99mTc] tilmanocept accurately detects sentinel lymph nodes and predicts node pathology status in patients with oral squamous cell carcinoma of the head and neck: results of a phase III multiinstitutional trial. Ann Surg Oncol 2015; 22:3708-15.
33. Wallace A, Han LK, Povoski SP, et al. Comparative evaluation of [99mTC] tilmanocept for sentinel lymph node mapping in breast cancer patients: results of two phase 3 trials. Ann Surg Oncol 2013; 20:2590-9.
34. Pelosi E, Bello M, Giors M, et al. Sentinel lymph node detection in patients with early stage breast cancer: comparison of periareolar and subdermal/peritumoral injection techniques. J Nucl Med 2004; 45:220-5.
35. Seradour B. Breast cancer screening in France: an overview in 2009. Rev Prat 2010; 60:191-9.
36. Francis A, Haugen C, Grimes LM, et al. Is sentinel lymph node dissection warranted for patients with a diagnosis of ductal carcinoma in situ. Ann Surg Oncol 2015; 22:4270-9.
37. Hieken T. The promise of axillary imaging in individualized surgical management of breast cancer patients: another step forward. Ann Surg Oncol 2014; 21:3369.
38. Caudle A, Kuerer H, Le-Petross H, Yang W, Yi M. Predicting the extent of nodal disease in early-stage breast cancer. Ann Surg Oncol 2014; 21:3440-7.
39. Mansel RE, Fallowfield L, Kissin M, et al. Randomized multicenter trial of sentinel lymph node biopsy versus standard axillary treatment in operable breast cancer: the ALMANAC trial. J Natl Cancer Inst 2006; 98:599-609.
40. Feldman S, Bansil H, Ascherman J, Grant R, Borden B. Single institution experience with lymphatic microsurgical preventive healing approach (LYMPHA) for the primary prevention of lymphedema. Ann Surg Oncol 2015; 22:3296-301.
41. DiSipio T, Rye S, Newman B, Hayes S. Incidence of unilateral lymphedema after breast cancer: a systematic review and meta-analysis. Lancet 2013; 14:500-15.
42. Vignes S, Porcher R, Arrault M, Dupay A. Long-term management of breast cancer-related lymphedema after intensive decongestive physiotherapy. Breast Cancer Res Treat 2007; 101:285-90.
43. Yah-Chen T, Ying XU, Cormier S, et al. Incidence, treatment costs, and complications of lymphedema after breast cancer among women of working age: a 2-year follow-up study. J Clin Oncol 2009; 27:2007-14.
44. 44. Klimberg VS. A new concept towards the prevention of lymphedema: axillary reverse mapping (ARM). J Surg Oncol 2008; 97:563-4.
45. Boccardo F, Casabona F, De Cian F. Lymphedema microsurgical preventive healing approach: a new technique for primary prevention of arm lymphedema after mastectomy. Ann Surg Oncol 2009; 16:703-8.
46. Namm J, Mueller J, Kocherginsky M, Kulkarni S. The utility of sentinel lymph node biopsy in patients with ductal carcinoma in situ suspicious for microinvasion on core biopsy. Ann Surg Oncol 2015; 22:59-65.

Corporate Governance Ratings on Listed Companies: An Institutional Perspective in Latin America

Guadalupe del Carmen Briano-Turrent [1, *], Lázaro Rodríguez-Ariz [2]

[1] Universidad Autónoma de San Luis Potosí, Sierra Leona No. 550, Lomas 4a, Sección, 78210 San Luis Potosí, Mexico

[2] Universidad de Granada, Campus Universitario de Cartuja, 18011 Granada, Spain

* Corresponding Author. E-mail: : guadalupe.briano@uaslp.mx

Abstract. The aim of this paper is to analyse whether institutional factors determine the level of corporate governance compliance among major listed companies in emerging markets of Latin America, a region characterized by a poor legal system, highly concentrated ownership structures, and capital markets relatively less developed. The paper used an unbalanced panel data consisting of 826 observations of the highest ranked companies on the stock exchange indices of Argentina, Brazil, Chile and Mexico during the period 2004–2010.

The results provide strong empirical evidence that board independence, ownership concentration and stakeholder orientation affect positively corporate governance ratings, while board size decreases corporate governance compliance in Latin American countries. The study fills a gap in the Latin American literature, providing useful information for determining policies on corporate governance and, in general, for managers and investors of listed companies in Latin America.

Keywords: Corporate governance; Ratings; Institutional theory; Emerging markets; Latin America.

1. Introduction

Corporate Governance [CG] is a relevant issue in academic writing and finance and accounting fields due to the chain of financial scandals around the world. CG monitors the effectiveness of management and ensures legal compliance by preventing irregular and improper behaviour. In this sense, leading global institutions such as the Organization for Economic Cooperation and Development [OECD], the International Finance Corporation [IFC] and the World Bank, strongly emphasize the development of different regulations, guidelines and good governance codes around the world. OECD affirms that CG "has implications for company behaviour towards employees, shareholders, customers and banks". A corporation's corporate governance structure is an important criterion when investors make investment decisions (Epps & Cereola, 2008). In the case of emerging markets, compliance with good CG practices is an effective substitute when legal environments and regulatory frameworks are weak and highly concentrated ownership structures predominate.

In this context, companies that improve their CG practices could be able of protecting shareholders rights and increase the confidence of investors (La Porta, López-de-Silanes, Shleifer, & Vishny, 2000). As a result, different ratings on CG [CGR] have been proposed by institutions and academics around the world. The construction of a rating or index is beneficial as it integrates the various elements of a firm's governance system into one number. Although there is no standardized system to measure the compliance on CG, prior research has been developed several CGR mostly for Anglo-Saxon and continental European countries (Gompers et al., 2003 and Klapper and Love, 2004). The main objective of the CGR is to assess and compare the companies' governance score regarding the accepted standards issued by regulatory bodies in a particular institutional context (Al-Malkawi, Pillai, & Bhatti, 2011).

Regarding to factors that affect the CG compliance, prior studies have recognized the institutional framework in emerging countries (Aguilera & Jackson, 2010). Institutional theory integrates a wider understanding related to cultural dimensions and formal factors of the firm in a modern society (Davis, 2005). Therefore, CGR may be notably influenced by institutional factors such as culture, legal

structures and financial markets (Creed et al., 2010, Peng et al., 2009 and Suddaby and Greenwood, 2005). On the other hand, agency theory points out the conflict of interest between management and owners due to separation of ownership and control. To minimize this divergence and reduce agency costs, this theoretical approach suggests the adoption of internal and external mechanisms of CG by companies (Haniffa and Hudaib, 2006 and Tariq and Abbas, 2013). In this study, institutional and agency theories are adopted as the main reference frameworks to empirically describe the factors that affect the CG compliance in Latin American listed companies. Formal factors at the macro level (legal system and government initiatives such as CG codes) and at inter-organizational level (board structure, ownership concentration or leverage) play an important role by adopting of CG practices (Boliari and Topyan, 2007 and Campbell, 2007).

Latin America is characterized by poorer CG and inferior legal system, highly concentrated ownership structures, and capital markets relatively less developed in comparison to more developed OECD economies (Blume & Alonso, 2007). The conflict of interest between major and minority shareholders reduces overall shareholder value and increases the expropriation of minority shareholders. Our motivation stems from the growing relevance of CG for investor confidence in the region and the absence of prior research in Latin America, which partly stems the scarcity of relevant data (Kabbach de Castro, Crespi-Cladera, & Aguilera, 2012). We pay attention in the institutional context, efficiency and legitimacy of CG mechanisms in an international business environment. In this sense, the question in this study is how institutional and agency theories could identify those formal and informal factors that promote CG compliance in Latin American listed companies?

We contribute to the literature in several ways. First, we propose a CGR which is based on the institutional and regulatory framework of the region. Second, we support our results using a sample of 826 non-financial firms in fourth largest stock exchanges of the Latin American region (Argentina, Brazil, Chile and Mexico), over the period 2004–2010. Third, we identify institutional formal and informal factors may be significant to CG compliance through GMM method addressing the reverse causality problem using suitable lagged values of the explanatory variables as instruments (Blundell and Bond, 1998 and Pindado et al., 2014). Finally, this study may provide useful information for determining policies on corporate governance and, in general, for managers and investors in listed companies. To the best of our knowledge, this is the first study that focuses on Latin America emerging countries combining the institutional and agency theories in a context characterized by a weak legal system and a lower shareholder protection.

The rest of the study is organized as follows. The authors, first presents a review of relevant literature and develop the study hypotheses. Secondly, the data and construction of the CGR are presented. Thirdly, we describe the data and methods of analysis. Fourthly we discuss the main results. Last section concludes.

2. Literature review

The compliance on CG can be looked upon from different theoretical perspectives, for instance economic, legal, social and applied finance (Ariff and Ratnatunga, 2008 and Tariq and Abbas, 2013). The theoretical foundation can be found in agency theory which points out that higher ownership concentration results in a conflict between majority and minority shareholders, with several well-known cases of expropriation (Jensen & Meckling, 1976). The major problem of this conflict is that minority shareholders are not protected against expropriation by majority and it is mainly due to weak legal structure (enforcement) of countries. Agency theoretical framework has tried to explain the relationship between shareholders and management, seeking the interest's alignment of managers and shareholders with CG mechanisms (Lopes & Walker, 2012). However, agency theory is limited and does not explain the multi-dimensional complexity and character of the CG phenomenon in an international business context (Adegbite, 2015). The conceptual framework of institutional theory is much broader and deeper than agency theory, since accounts for the deeper and resilient aspects of socio-cultural structure, and integrates the process by which organizational schemas, rules, norms, and routines are established as guidelines for corporate behaviour (Scott, 2004). Furthermore, this theoretical approach is most suitable to explain CG practices in contexts characterized by small stock

market, a higher level of ownership concentration in the hands of a few shareholders, and a strong link between CG structures and institutional development (Baixauli-Soler & Sanchez-Marin, 2011). Globerman and Shapiro (2003) observed that formal institutions – regulation, financial markets, transparency and accountability – strengthen the governance structure and attract more foreign investment. However, in these countries, informal institutions play an important role when formal mechanisms prove to be inadequate (Estrin & Prevezer, 2011). The above has caused an increase in the adoption of good governance practices as part of firms' strategy to increase investors' confidence. In the case of emerging markets, the institutional conditions may explain variations in the level of business activity and corporate practices (De Clercq, Danis, & Dakhli, 2010).

The adoption of corporate practices and principles co-evolving with institutions might become institutionalized. Institutionalization implies a certain degree of internalization and cognitive belief in the practice which is quite distinct from decoupling practices (Terjesen, Aguilera, & Lorenz, 2015). The Latin American model of CG is characterized by undeveloped capital markets, weak institutional environments, highly concentrated ownership structures, and lower protection of investors (Chong and López-de-Silanes, 2007 and Djankov et al., 2008). The proliferation of governance codes and adoption of best practices in Latin America – especially in the larger economies of Argentina, Brazil, Chile and Mexico which capture 70% of regional market capitalization (S&P, 2010) – and the creation of institutions like the Latin American Corporate Governance Roundtable as a joint initiative of the International Monetary Fund [IMF], the World Bank, and state and private actors from Latin and OECD countries, promotes a new era on CG in the region (Diamandis & Drakos, 2011). The guidelines issued by the OECD, codes of good governance and the regulations issued in each of these countries have all contributed to raising the CG compliance in issues related to the board of directors, shareholder rights, conflicts of interest, ownership structure and support committees of the board.

Latin American countries have adopted voluntary practices of CG to cover for the limitations of the regulatory framework. Good governance codes and laws prevailing in the region have been based on the "White Paper" and GC principles of the OECD. These CG codes have promoted transparency and market efficiency, the protection of shareholders and effective board of directors monitoring. Particularly, Argentina, Brazil and Mexico have opted for soft laws, through the principle of "comply or explain". Chile has focused on hard laws and legal enforcement which aim to strengthen the board of directors, auditing committee functions, shareholders rights and reduce conflicts of interest, however there are inefficient self-regulation practices concerning the capital markets (Lefort & González, 2008).

Recent studies have adopted the institutional-agency theories to analyse the factors that influence on CG compliance in different contexts. For instance, Seal (2006) proposes an institutional theory of agency, which may defined as the analysis of managerial behaviour in giant, widely owned corporations where managerial action is influenced by institutionalized practices that affect corporate practices and performance. This combination of theories establishes managerial behaviour has been influenced and legitimized by the dominant discourse of CG – the agency theory. Institutional theory is defined as a set of formal and informal rules that affect business activity (North, 2005). In this regard, both formal (e.g. government initiatives, laws; Campbell, 2007) and informal institutions (e.g. corporate culture and strategy; Boliari & Topyan, 2007) are regarded as antecedents to action by defining the CG practices. According to DiMaggio and Powell (1983), institutional theory indicates that firms tend to incorporate external norms and rules into their operations and structures in order to gain legitimacy and social acceptance. Thus, it can be argued that companies may gain acceptance and legitimize their operations by engaging in CG compliance (Ntim, Lindop, & Thomas, 2013). Thereby, all forms of institutions that manage human interactions via cognitive, normative, and regulative processes influence organizational decision-making (Trevino, Thomas, & Cullen, 2008).

3. Hypothesis development

Institutional theory emphasizes that legal rules and norms form an important element of national institutional systems (Filatotchev, Jackson, & Nakajima, 2013). The groundbreaking work by La Porta, López-de-Silanes, Schleifer, and Vishny (1998) argue that a common element in differences

between countries is the degree of investor protection against abuses by the management team and majority shareholders. The degree of law enforcement creates cross-country differences. For instance, civil laws give investors weaker legal rights than common laws do. The difference in legal protections of investors might help explain why firms are financed and owned so differently in different regions. Shleifer and Vishny (1997) conclude that a very high ownership concentration may be a reflection of poor investor protection. Differences in legal systems have implications for transparency on CG practices, directly or indirectly, with firms in common-law countries disclosing more CG information than those located in civil-law countries (Li & Moosa, 2015). Institutional theory argues that firms tend to incorporate external norms and rules in order to gain confidence and legitimacy in the market (Scott, 1987). Various studies have sought to measure the degree of enforcement, and Leuz, Nanda, and Wysocki (2003) suggested it can be measured through three variables: (a) the efficiency of the judicial system; (b) an evaluation of the rule of law; (c) an index of corruption. Kaufmann, Kraay, and Mastruzzi (2011) proposed a series of governance indicators including the dimensions of regulatory quality, the rule of law and the control of corruption. The Worldwide Governance Index [WGI], on the other hand, published by the WGI (2014), includes six dimensions for 213 economies, assessed for the period 1996–2010: (1) accountability; (2) political stability and absence of violence; (3) governmental effectiveness; (4) regulatory quality; (5) the rule of law; (6) the control of corruption. Hence, our first hypothesis is that:

Hypothesis 1.

There is a positive relation between the WGI and the CGR in Latin American countries.

The dimensions of CG contained in the codes of good governance and regulatory framework constitutes formal factors that may influence managerial decisions related to the compliance level on CG (Ho & Wong, 2001).

- Size of the board: The board should comprise a reasonable number of directors; its size directly affects its functioning and supervisory capacity (Gandía, 2008). Larger boards enjoy greater diversity and tend to have more experienced members, which affects the CGR (Gallego Álvarez et al., 2009 and Laksamana, 2008). Various studies corroborate the presence of a positive relationship between board size and the level of CG compliance (Barako, Hancock, & Izan, 2006; Hussainey & Al-Najjar, 2011).

Hypothesis 2a.

Board size has a positive impact on CGR in Latin America.

- Composition of the board: External (non-executive) directors are not part of the company's management team and so are in a better position to monitor management performance (Donnelly & Mulcahy, 2008). They have an added incentive to facilitate supervision by shareholders because their own reputation depends on the corporate performance (Fama & Jensen, 1983); moreover, they are the most effective agents for maximizing shareholder value (Rouf, 2011). Most studies affirm there is a positive relationship between the independence of the board and CGR (Abdelsalam and Street, 2007, Kent and Stewart, 2008 and Samaha and Dahawy, 2011).

Hypothesis 2b.

There is a positive relationship between the proportion of independent directors and CGR in Latin American countries.

- COB-CEO duality: COB-CEO duality refers to the situation in which the same person holds both positions in a company. According to Haniffa and Cooke (2002), separation between the two positions helps improve the quality of supervision and reduces the advantages gained by withholding information, while the concentration of power is associated with reduced transparency and lower quality of CG information (Laksamana, 2008).

Hypothesis 2c.

There is a negative relationship between COB-CEO duality and CGR in Latin America.

- The presence of women on the board: In recent years the issue of gender diversity in business has received considerable research attention. Women provide viewpoints, experiences and work styles that differ from those of their male counterparts (Torchia, Calabró, & Huse, 2011). Among the variables that have been associated with the presence of women on the board is the level of CG

transparency, thus increasing the board's capacity to supervise the process of CG transparency and compliance (Gul, Srinidhi, & Ng, 2011). Several studies have suggested that gender diversity is associated with a higher quality of boardroom debate and more effective communication (Hillman, Shropshire, & Cannella, 2007; Huse & Solberg, 2006), thus facilitating greater availability of information to investors.

Hypothesis 2d.

The proportion of women on the board is positively associated with the CGR in Latin America.

- Ownership structure: An important factor shaping the CG system is the company's ownership structure, defined as the degree of concentration that determines the distribution of power and corporate control, or as the proportion of voting shares owned directly or indirectly by senior management, board members or their relatives (Owusu-Ansah, 1998). When the ownership structure is diffuse, greater supervision is needed in order to maintain fair access for minority shareholders. Companies with widely dispersed ownership tend to disclose more information in order to reduce the costs of control by shareholders (Haniffa & Cooke, 2002). Furthermore, this transparency on CG is increased when there are external shareholders (Donnelly & Mulcahy, 2008). Some studies have reported a negative relationship between ownership concentration and the level of information disclosure (Barako et al., 2006, Gandía, 2008 and Vander Bauwhede and Willekens, 2008, among others). In companies with large individual shareholders or a high concentration of ownership, information is transferred directly through informal channels, or there may simply be a greater alignment of interests, thus reducing the need to make information public.

Hypothesis 2e.

There is a negative association between ownership concentration and CGR in Latin American countries.

- Family-controlled firms: According to the agency theory, family controlled firms create agency costs. The risk of wealth expropriation from minority shareholders is higher when ownership is concentrated and held by family members (Barontini & Caprio, 2006). In this sense, CG compliance in family firms may become inconsistent with wealth maximization. The combination of ownership and control in family firms could generate and excessive role by the owner through its leadership, which could lead to problems of management entrenchment. Faccio and Lang (2001) argue that family firms present a poor performance compared to no family firms, while San Martin-Reyna and Duran-Encalada (2012), anticipate problems associated with family firms and composition of directors. Family owners could favour family interests over the firm's interests (e.g. minority shareholders) and have incentives to be engaged in opportunistic behaviours, because of loyalty towards the family (Schulze, Lubatkin, Dino, & Buchholtz, 2001). Thus we set the following hypothesis.

Hypothesis 2f.

Family-controlled firms obtain a lower corporate governance rating (CGR) than non-family firms in Latin American listed firms.

Institutional theory suggests that a firm's right to exist is legitimized if its value system is consistent with that of the larger social system of which it is part of, but threatened when there is actual or potential conflict between the two value systems (Suchman, 1995). Diverse interest groups influence decision-making and the values adopted by the firm (Donaldson & Preston, 1995). Bradley, Schipani, Sundaram, and Walsh (1999) identified two types of culture in companies: community or stakeholder-oriented culture, with a broad range of members having a legitimate interest in corporate activities, and shareholder-oriented culture, with a contractual outlook, in which companies are viewed as tools for creating shareholder value, and in which other stakeholders have less legitimacy and influence over management. In line with Simnett, Vanstraelen, and Fong Chua (2009), in this study we consider the stakeholder vs. shareholder orientation as a dimension of organizational culture. Under this approach, Smith, Adhikari, and Tondkar (2005) revealed that companies with a stakeholder-oriented approach disclose more information as part of their strategic management approach in order to strengthen relations with stakeholders, while Basu and Palazzo (2008) suggest that companies could improve the credibility of their communication by exposing transparency to

questioning through stakeholders. In the same line, Jansson (2005) argues that the stakeholder orientation depends of the governance and ownership structure of the firm and the legal environment.

Hypothesis 3.

The CGR is higher in firms with a stakeholder orientation than in those oriented towards shareholders.

Research on the relationship between CG transparency and innovation has been limited (Miozzo & Dewick, 2002). According to O'Sullivan (2000, p. 1), innovation is performed with the aim of increasing product quality and/or lower production costs. Innovation can provide the critical component of a firm's competitive strategy. Gill (2008) found that those companies that follow innovation as a strategy disclose more information to signal commitment to the project, potentially inducing a rival's exit. Inside directors are generally associated to innovative strategies, because they have a better knowledge of the company. As a consequence, detailed information is required to make effective strategic decisions and monitoring (Zahra, 1996). Competitive pressure might be alleviated when firms that innovate disclose more corporate information to induce rivals to wait and imitate instead of simultaneously invest in innovation (Pacheco-De-Almeida & Zemsky, 2012). By considering than Latin American countries are characterized by a higher inside director's rate (Black, Gledson de Carvalho, & Gorga, 2010), we argue that innovation strategy influences the CGR leading to the following hypothesis:

Hypothesis 4.

Firms with an innovation strategy obtain higher CGR than firms with a no innovation strategy.

4. Study methodology

The object of this study is to analyse the CG ratings of those major listed companies in Argentina, Brazil, Chile and Mexico. For the sample of firms in these four countries, we selected the most representative of each country. According to Kitagawa and Ribeiro (2009) the purposes of analysis, we excluded those in the banking and insurance sectors, because these are more strictly regulated and are subject to greater scrutiny in terms of corporate information disclosure (Garay & González, 2008). The information needed to construct the index of CG and the set of explanatory variables used was obtained from the annual reports and websites of the selected companies, by means of content analysis. The content analysis could be used to identify the different CG categories as reported by sample firms to distinguish the different levels of compliance, depending on the nature of its business and global environment. Given the qualitative nature of CG disclosure, we perform a content analysis focusing on the volume and intensity of disclosure using the number of words and sentences with to different items of CG categories and sub-categories in order to integrate the CGR (Lajili & Zéghal, 2005).

For clustering purposes, the companies were ranked according to the Global Industry Classification Standards [GICS], which are widely accepted in the business and academic worlds (Bhojraj, Lee, & Oler, 2003). Outliers, or extreme values, for the financial variables were identified and analysed, and values above the 99th percentile were assigned the value of this percentile. Values below the first percentile for each variable were truncated in the same way (Braga-Alves & Shastri, 2011).

Initially, 155 companies were considered, but 20 belonging to the financial sector were excluded as were a further seven for which there was insufficient information for analysis. Thus, the final study sample was constituted of 128 companies. Regarding the number of observations included in this empirical study covering the time period from 2004 to 2010, data were obtained for 101 companies in 2004, 111 in 2005, 116 in 2006, 123 in 2007, and 125 in 2008, 2009 and 2010. A total of 826 observations were obtained for the whole period of analysis. Table 1 shows the composition of the study sample by country. The predominant sectors in these countries are related to materials, consumer staples and utilities.

Corporate governance rating

Several indices on CG have been developed for Anglo-Saxon and continental European countries (Gompers et al., 2003 and Klapper and Love, 2004). According with institutional theory, the legal and

institutional context of each country is a key factor in the selection of the elements of an index (Hossain & Hammami, 2009). This study proposes a CGR that evidently reflects the nature of emerging Latin American institutional framework, using a combination of information required by the rules and codes of good governance in the selected countries. For instance, codes of good and regulatory framework in Argentina, Brazil and Mexico and Chile. In this study, we glean support for the index from the OECD principles, the codes of good governance in each country, and previous studies in the region.

The overall CGR composed by 43 items, with a maximum value of 100, was obtained by summing four sub-indices: (1) composition and performance of the board, (2) shareholders rights, (3) ethics and conflicts of interest and (4) other information related with CG. In compiling the overall index, each sub-index is weighted as 53, 18, 16 and 13%, respectively (Lefort & González, 2008). Each sub-index was in turn comprised of a series of factors with the same weights (for more detail see Table 2). The composition and performance of the board sub-index captures board independence, mission, functions, structure and effectiveness. Autonomy is established through various factors of board independence, including the COB-CEO duality and the presence of support committees (nominating, remuneration, corporate governance, auditing). Furthermore this sub-index also contains measures of board remuneration, selection, removal or re-election procedures, and disclosure of profile or curriculum of directors including the document that establishes the norms of conduct for the board members. However, most of the items in this category (at least 14 out of 24) are allocated to measures that reflect board independence.

Shareholders rights comprise the second sub-index, the purpose of which is to identify the mechanisms that encourage the alignment between board of directors and managers interests with those of shareholders. For instance, description of shareholding voting process, pyramidal structures that reduce the concentration of control, information of the agenda, shareholders agreements and resolutions proposed for its adoption. The sub-index related to ethics and conflicts of interest attempts to measure conflicts of interests and related party transactions, company operations with its directors and managers, significant transactions between the company and significant shareholders and ownership composition. The final sub-index deals with other related information with CG. It attempts to measure a company's public commitment with good corporate practices. The use of international accounting principles, the services of a recognized auditing firm, sanctions against the management for breach of their CG practices, financial performance disclosure, and practices of good governance, score well in this category. The index allows each element to be equally important and does not distinguish subjective selection of the most influential characteristics (Berglöf & Pajuste, 2005). Nevertheless, we compute a weighted sum of the four dimensions in our calculations.

Model specification and measurement of the variables

The following multiple regression model was applied to test our hypotheses. The dependent variable is the proposed CGR. The independent and control variables were determined on the basis of previous studies and are detailed in Table 3.

Control variables

Leverage. Companies with higher debt levels are generally under closer scrutiny by creditors, and have greater incentives to disclose more information about their management performance (Samaha et al., 2012 and Xiao et al., 2004).

Age of the company. The age of the firm can influence the level of corporate transparency, as this represents the company's stage of development and growth (Owusu-Ansah, 1998). Under this premise, younger firms tend to disclose less information than more mature ones, for three reasons: (1) greater transparency can affect their competitive advantage; (2) the cost and ease of information processing and disclosure is greater; (3) the relative absence of such information. In our study, we expect to find a negative relationship between these two variables (Hossain & Hammami, 2009).

Size of the company. Most studies have found that company size positively affects the level of corporate information disclosure (Bassett, Koh, & Tutticci, 2007). Larger companies have certain characteristics that differentiate them from smaller ones, such as the greater diversity of products,

more complex distribution networks and greater need for funding from capital markets (Gallego Álvarez et al., 2009).

Profitability. Managers disclose more detailed information to ensure the continuity of their positions and remuneration and as a sign of institutional confidence. Inchausti (1997) argues that more profitable companies make greater use of information in order to obtain a competitive advantage, while firms with poor performance may be less transparent. Previous studies mainly reflect a positive relation (Apostolos & Konstantinos, 2009).

Business sector. The business sector is another variable that has often been used to account for the amount of information provided by companies (Eng & Mak, 2003). Companies operating in the same sector are believed to disclose similar information in the market, to avoid sending a bad signal to investors (Watts & Zimmerman, 1986). Companies operating in more politically visible sectors have greater incentives to voluntarily disclose information in order to minimize any political costs (Collet & Hrasky, 2005). The studies that have reported a significant relationship between the business sector and the disclosure of information include Gandía (2008), Bonsón and Escobar (2006) and Nagar, Nanda, and Wysocki (2003).

5. Results

Descriptive analysis of the data

Table 4 and Table 5 summarize the descriptive data for the dependent and independent variables for 2004 and 2010 study periods.

In the countries analysed, the CGR increased during the study period; the average value was 0.36 (median 0.33) in 2004 in Argentina, while in Brazil it was 0.48 (median 0.49). For firms in Chile, the average was 0.53 (median 0.50), and in Mexico, 0.66 (median 0.67). Mexico presented the highest index value, followed by Chile, Brazil and Argentina. In 2010 the index showed an increase in the four countries under analysis. For instance, Argentina averaged 0.64 (median 0.66), Brazil 0.72 (median 0.74), Chile 0.64 (median 0.61), and in Mexico, 0.78 (median 0.79). The results suggest a favourable evolution of the formal institutional environment in the region, since the codes of good governance and regulations have increased and revised several times. Differences between countries are mainly due to the codes of good governance of each country, which require of different levels and dimensions of corporate transparency.

Regarding the institutional formal factors, legal system recorded an average value of 0.59 for the WGI, with Chile obtaining the highest value, followed by Brazil, Mexico and Argentina. The explanatory variables related to CG dimensions include the size and composition of the board. The firms analysed had an average of 9.6 directors in 2004 and 10 in 2010. Although there was no significant variation in the average board size during the study period, we did find that the rules and codes of good governance within each country applied diverse criteria regarding this parameter (recommending 5–9 members in Brazil, a minimum of 7 in Chile, and between 3 and 15 in Mexico, with no recommendation being made in Argentina). Regarding board composition, for the region as a whole the average number of external directors was 0.33 in 2004 and 0.38 in 2010. In Argentina, the corresponding values were 0.20 (median 0.18) in 2004 and 0.31 (median 0.32) in 2010. In Brazil, these values were 0.25 (median 0.21) in 2004 and 0.34 (median 0.33) in 2010. According to Black et al. (2010), the independence of the board is a notoriously weak area in Brazil, with most company boards being composed of representatives of the controlling group. Chile presented an average value for independent directors of 0.38 (median 0.33) in 2004 and 0.38 (median 0.33) in 2010. Finally, Mexican firms had an average of 0.46 (median 0.47) in 2004 and 0.49 (median 0.50) in 2010, and so these companies had the highest proportion of independent directors in our study group, perhaps because the code of corporate governance in this country stipulates a minimum proportion of independent directors (25%), whereas the other countries specify neither their number nor their proportion.

COB-CEO duality was found in 28.7% of the firms analysed in 2004 and in 21.6% in 2010, occurring most frequently in Argentina and Mexico. In these emerging economies in Latin America, there is a growing presence of women on boards of directors. Nevertheless, the total numbers remain

far from significant; the highest level of female participation is found in Brazil, but the regional average was barely 4% in 2004 and 5% in 2010.

A clear ownership concentration is observed in the region, although this has been declining in the analysed period. Among the companies examined, the average proportion of shares held by the top ten shareholders was 57% (median 51%) in 2004 and 55% (median 54%) in 2010. With respect to family-controlled variable, we observe that 41.6% of the firms are controlled by families, compared with 58.4% of non-family firms. In 2010, the percentage of family firms increased to 47.2% while non-family companies decreased to 52.8%.

With regard to informal factors, we observed an institutionalization of corporate culture towards a stakeholder orientation. Thus, only 51.5% of these companies were basically stakeholder oriented in 2004, while in 2010 this figure had increased to 83.2%. This trend reflects growing interest among companies in considering a broader range of participants, and this in turn has a bearing on levels of corporate transparency. In respect of the corporate strategy followed by studied firms, predominates innovation strategy (77.2%) compared with no innovation strategy (22.8%) in 2004, while in 2010 the adoption of innovation strategy increased to 82.4% compared with no innovation strategy (17.6%).

Regarding the control variables, the average level of leverage in the region was 23% in 2004 and 29% in 2010, and higher among companies in Brazil, Chile and Mexico. The average age of these firms, from their founding, was 48.31 years in 2004 and 49.95 years in 2010. Firm size, measured by the natural logarithm of its assets, was highest in Brazil, followed by Mexico and Chile. The regional average was 7.83 in 2004 and 8.72 in 2010. Finally, the descriptive statistics for the variable measuring financial performance [ROA] showed that the best performance was obtained in Argentina, followed by Mexico, Brazil and Chile. The overall average for these countries was 0.11 in 2004 and 0.10 in 2010.

Bivariate analysis (correlation matrix)

The potential multicollinearity among the explanatory variables was analysed to obtain the variance inflation factor [VIF] and level of tolerance. Table 6 (Panel A) shows the Pearson coefficients for all the study variables. This correlation analysis shows that the CGR is positively and significantly correlated with board size, board independence, COB-CEO duality, stakeholder orientation, size of the firm and year of study ($p < 0.01$, two-tailed test); and leverage ($p < 0.05$, two-tailed test).

The CGR is negatively correlated with the WGI and ownership concentration ($p < 0.01$, two-tailed test); and corporate strategy and industry type ($p < 0.05$, two-tailed test). We also observe that the highest value of the correlation between independent variables and CGR is 0.400 (board size). According to Gujarati (2003) correlations between the independent variables are not considered harmful to the multivariate analysis at least exceeding 0.80.

Panel B shows the coefficients for VIF and tolerance, which must be within the limits proposed by Xiao et al. (2004), i.e. less than 2 for VIF and above 0.60 for the tolerance level. In this line, Neter, Wasserman, and Kutner (1989) proposed that the VIF coefficient should not exceed 10, since that would indicate the presence of damaging multicollinearity. On the other hand, if the average VIF were substantially less than 1 this would indicate that the regression analysis might be biased (Bowerman & O'Connell, 1990). Our study obtained an average VIF of 1.228, which is in line with the values obtained by Hossain and Hammami (2009) and Shan and Mclver (2011), who confirmed that their model had no multicollinearity, with VIF values of 1.47 and 1.42 respectively. The VIF was within the recommended limits, while the correlation matrix revealed no major correlation problems among the variables.

Analysis of results

Table 7 shows the multivariate analysis results for the proposed hypotheses. First, multiple regression analysis [OLS] with robust estimator (VCE) was performed, including the industry and the year of study as dummy variables, to incorporate their possible effects. Subsequently, we use the GMM system to account for endogeneity of all time-varying explanatory variables (Bloom and van Reenen, 2007 and Pindado et al., 2014). We have adopted GMM to control for endogeneity and reduce the risk of obtaining biased results due to correlation between error term and explanatory

variables. GMM relies on set of "internal" instruments (lags of explanatory variables), eliminating the need of external instrumental variables (Wintoki, Linck, & Netter, 2012). The multiple regression model (Model 1) was found statistically significant ($p > 0.000$). The adjusted coefficient of determination (R2) indicates that 49.35% of the variation in the dependent variable is explained by the independent variables. The coefficients show that statistically significant formal institutional variables in the model are legal system (+), size of the board (+), independence of the board (+), and participation of women on the board (−). The significant informal institutional variables in the model are stakeholder orientation (+) and the innovation strategy (+). Finally, COB -CEO duality and ownership concentration are not significant in this model. Regarding the control variables, leverage (+), age of the company (−) and its size (+) are significant, while profitability does not present any association.

Continuing our analysis, the panel data highlighted problems of heteroscedasticity, endogeneity and correlation. To address these issues, we estimate the model using GMM method, because it is an instrumental variable estimator that embeds all other instrumental variables as special cases (Pindado et al., 2014). Model 2 shows that the positive, statistically significant relations were independence of the board ($p = 0.05$), ownership concentration ($p = 0.05$), stakeholder orientation ($p = 0.01$) and age of the company ($p = 0.10$). We could identify a negative relation between CGR and board size ($p = 0.05$). No significant relationships were found between CGR and legal system, COB-CEO duality, gender, family controlled firm, strategy, leverage, company size and profitability.

The results obtained also suggest that CGR in the energy sector is higher than in other sectors. From 2005, significant differences began to appear in the levels of transparency in the countries analysed. These findings are supported by Archambault and Archambault (2003), who concluded that the decision to adopt and publish corporate information is influenced by informal factors such as culture, regulatory system and the corporate system. Berglöf and Pajuste (2005), on the other hand, suggested that companies' corporate practices depend on the legal environment and practices prevailing in each country, company size and the concentration of ownership. Barako et al. (2006) carried out a longitudinal study using panel data methodology, and found a significant association between the level of corporate transparency and the CG attributes of the company, such as ownership structure and other characteristics. In the same vein, Samaha et al. (2012), studying firms in Egypt, argued that the proportion of independent directors and firm size are two factors that positively affect the level of CG compliance.

In this study, we observed a significant inverse association between the level of CG transparency and independence on the board, and so hypothesis H2a, which predicted a positive relationship between these variables, is rejected. These results are in line with those found by Mak and Kusnadi (2005), who argue that larger boards inhibit the motivation and participation of their members in the strategic taking decisions process, and therefore its impact negatively on the CG compliance.

The results for the CG dimensions show there is a positive and statistically significant (5% level) between independence of their members with CGR; therefore, hypotheses H2b is accepted. These results are aligned with those reported by Samaha et al. (2012) who showed that CGR for listed companies in Egypt increases in proportion with the number of independent directors and company size. In this context, our study shows that diversity in the boards provide the experience and knowledge necessary for adequate performance of their functions, and tend to increase the level of CGR (Ezat and El-Masry, 2008, Gandía, 2008, Kent and Stewart, 2008 and Willekens et al., 2005). The presence of independent members on the board represents a means of control that improves its effectiveness, focusing its attention on the actions of the management team and on ensuring the shareholders' goals are achieved, all of which is reflected in a higher level of CGR (Fama & Jensen, 1983). These consequences have also been reported for the level of CG transparency in other emerging countries (Ezat and El-Masry, 2008 and Samaha and Dahawy, 2011).

With respect to ownership concentration, the results show a significant and positive influence on CGR, opposite our established premise. In this sense we have rejected the H2e which suggested a negative relation. According to Haniffa and Cooke (2002), the ownership concentration could reduce the freedom of the management team and lead to a more efficient behaviour such as CG compliance.

Hypothesis 3 regarding to stakeholder orientation has been accepted. Thus, firms with a stakeholder-oriented approach will tend to adopt more CG practices as part of their strategic management and as a process of continuous and interactive communication between the company and its stakeholders (Fasterling, 2012). The significant control variables in the model is the age of the company, with a positive impact on the CGR, a finding that is in line with the results of Hossain and Hammami (2009) and Owusu-Ansah (1998), who argued that younger firms publish less information to maintain their competitive advantage, and also because they have a shorter history to communicate.

Regarding the variables that were not statistically significant, we observed that the legal system is not a determinant factor in CGR in Latin American region. Jaggi and Low (2000) considered the legal system to be the most significant institution affecting business activity, while Bushman, Piotroski, and Smith (2004) observed a positive relationship between the level of CG compliance and the strength of the legal system. However, in Latin American case this variable is not significant.

The hypotheses rejected concerned the COB-CEO duality (H2c), gender of the board (H2d), the family controlled firm (H2f), strategy (H4), and the control variables of leverage, size and profitability, none of which accounted for the CGR. Previous studies have suggested that in the case of emerging or developing countries the results of this type of analysis could differ from those found in developed economies (Archambault & Archambault, 2003). On the other hand, COB-CEO duality and family controlled firms do not seem to affect the level of CG compliance, which is in line with the results obtained by Ho and Wong (2001), Eng and Mak (2003) and Haniffa and Cooke (2002).

6. Conclusions

This study contributes to the literature pertaining to how formal and informal factors promote a higher CG compliance on listed companies of Latin American emerging markets. Most of prior research has focused on developed countries and they have associated institutional factors such as culture, legal system and financial factors with corporate and transparency practices, while agency theory suggest that internal and external dimensions of CG (board of directors, ownership concentration, legal system) could minimize the conflict of interest between majority and minority shareholders in countries where legal system and shareholders' protection is poor (Creed et al., 2010, Peng et al., 2009 and Tariq and Abbas, 2013).

This study obtained a comparative study of four emerging economies in Latin America (Argentina, Brazil, Chile and Mexico), and the results indicate a rising trend in CG compliance during the period 2004–2010. Our analysis shows that the variables that affect CGR in this region are the independence of the board, the ownership concentration, stakeholder orientation, and the age of the company. These results are consistent with the findings of Kent and Stewart (2008) and Samaha and Dahawy (2011) who affirm that independent directors promote a higher supervision and control in order to keep their reputation on the market. By contrast, larger boards influence negatively in the level on CG because a greater diversity of opinions may hinder a consensus to adopt corporate practices. Contrary to the proposed hypothesis, our results demonstrated that ownership concentration affect positively the CGR, since the region is characterized by a weak legal system and a poorer protection of shareholders, so ownership concentration becomes a control mechanism to substitute this absence and promotes the adoption of good governance practices (Gandía, 2008 and Vander Bauwhede and Willekens, 2008). In the case of informal factors, stakeholder orientation motivates the credibility and transparency on corporate practices in weak legal environments (Basu & Palazzo, 2008). In the case of female presence in the board, family element and legal system were no significant in the analysis, since the institutional framework in Latin America differ from other countries previously studied, both normative and cultural aspects.

Our research also considered the endogeneity problem in the empirical analysis of CGR. The endogeneity problem in this issue is important, for it is highly likely that observable and unobservable institutional factors may affect CG compliance, and some of firm-specific characteristics could influence the rating on CG.

This study has the following limitations. First, the study variables, compiled from the companies' annual reports, inevitably reflected the subjective judgement of the researchers, which could lead to errors of interpretation and of information compilation. Second, the proposed CGR was un-weighted. On the one hand, this presents the disadvantage that all the index items are awarded the same importance; however, to the best of our knowledge there is no established methodology to assign a single weighting criterion for such an analysis, and the use of an un-weighted index does reduce the problem of subjectivity. Third, we focused on obtaining information on CG and the study variables from three main sources: the companies' annual reports, the CG reports and the companies' websites; thus we did not consider press reports or other communiqués that may be issued by listed companies. Fourth, there is some subjectivity in the selection of the explanatory variables. Given the extensive literature in this field of research and the large number of variables that have been identified, we chose to include those appearing most frequently in previous studies. This limitation is mitigated by the use of panel data methodology and GMM system, which takes into account the problem of omitted variables. Fifth, our study considers only the listed companies with the highest rankings in four Latin American stock markets, and omits companies in other indices and non-listed companies. Nevertheless, we achieved a sufficient number of observations for panel data analysis to be applied.

Despite the above limitations, the results obtained constitute a benchmark for managers responsible for determining CG policies and legislation in the countries under study. These results reflect the current status of the region concerning CG and could help identify the dimensions and elements of CGR that favour regional convergence. Moreover, this paper opens up interesting areas for future research, highlighting the impact of certain ethical values and risk taking behaviour. It would be useful to extend this study to consider the influence of other cultural variables such as the nationality, age, education and experience of company directors or managers on the level of CG compliance, and to analyse other formal institutional factors such as directors' remuneration, the composition of support committees and the frequency of board and support committee meetings. Another area of interest would be to analyse the impact of the rating and its sub-indices on measures of corporate performance or risk, or perhaps the impact of formal and informal institutional factors on these same measures. Furthermore, the study sample could be expanded to include companies listed on the continuous market. An important extension of our work would be to include other countries, both emerging and Anglo-Saxon and Continental countries, and thus incorporate a greater number of country-level variables. Another interesting line would be to carry out a longitudinal study by business sector, to determine the impact of formal and informal institutional factors on a specific area of activity, and at the same time, a larger number of countries might be included.

References

1. Abdelsalam, O., & Street, D. (2007). Corporate governance and the timeliness of corporate internet reporting by U.K. listed companies. Journal of International Accounting, Auditing and Taxation, 16(2), 111–130.
2. Adegbite, E. (2015). Good corporate governance in Nigeria: Antecedents, propositions and peculiarities. International Business Review, 24(2), 319–330.
3. Aguilera, R., & Jackson, G. (2010). Comparative and international corporate governance. The Academy of Management Annals, 4(1), 485–556.
4. Al-Malkawi, H., Pillai, R., & Bhatti, M. (2011). Corporate governance practices in emerging markets: The case of GCC countries. Economic Modelling, 38, 133–141.
5. Apostolos, K., & Konstantinos, A. (2009). Voluntary accounting disclosure and corporate governance: Evidence from Greek listed firms. International Journal of Accounting and Finance, 1(4), 395–414.
6. Archambault, J., & Archambault, M. (2003). A multinational test of determinants of corporate disclosure. The International Journal of Accounting, 38(2), 173–194.
7. Ariff, M., & Ratnatunga, J. (2008). Do accounting and finance tools serve governance? International Journal of Economics and Management, 2(1), 1–27.
8. Baixauli-Soler, J., & Sanchez-Marin, G. (2011). Organizational governance and TMT pay level adjustment. Journal of Business Research, 64(8), 862–870.
9. Barako, D., Hancock, P., & Izan, H. (2006). Factors influencing voluntary corporate disclosure by Kenyan companies. Corporate Governance: An International Review, 14(2), 107–125.

Transforming Society by Transforming Technology: The Science and Politics of Participatory Design

Peter M. Asaro [1, *]

[1] Departments of Philosophy and Computer Science, Beckman Institute, University of Illinois at Urbana-Champaign, 105 Gregory Hall, 810 S Wright Street, Urbana, IL 61801, USA

* Corresponding Author. E-mail: asaro@uiuc.edu

Abstract. This article attempts to shed historical light on some of the social, political, and ethical issues that have arisen from two disparate perspectives on technology which have both come to integrate an explicit consideration of social factors into systems design. It presents two distinct historical traditions which have contributed to the current field of participatory design methodologies—Joint Application Design (JAD®), and the British "socio-technical systems" and Scandinavian "collective resources" approaches—and which in practice integrated the end-users in different ways consequent upon their differing perspectives on workers, professional relationships to technology, and stated goals. One interest in examining the independent development of methodologies from these two perspectives is that, despite their differences, the approaches ultimately converged on a set of shared concerns and very similar practices.

The paper also examines the relation of these traditions to transformations in the theorization of business organization and trends of corporate restructuring which helped to secure a place for variants of related methodologies in major US and multinational corporations. It concludes with an examination of some broader issues in the relationship between technology and society and the prospects for the critical study of technology. I argue that participatory design and its related methodologies are best understood as a model for involving users, designers and the technology itself in a process of technological development. Rather than seeing participatory design as merely the insertion of public dialog within technological design practices, as several observers have done, we should see it as a model for the critical practice of developing technological designs.

Keywords: Participatory design; Critical theory; History; Technology; System design.

1. Introduction

While technological "progress" is not without many vocal and compelling critics, the fact that technology permeates our society is undeniable. The insight into the deep connection between the form of technology and the form of human life, so eloquently expressed by Heidegger in the quote above, has led to some very interesting work in the design of socio-technical systems. This article attempts to come to terms with some of the social, political, and ethical issues that arise from this work in a single, though broad, domain of socio-technological development—information systems design. I have chosen this particular domain because of an interesting convergence which has occurred in the practices which involve users in systems design.

There are almost as many different ways to design information systems as there are information systems, but there are also identifiable commonalities among groups in the various approaches to design. A good history and examination of the philosophical assumptions of different approaches to the design process is presented by Hirschheim, Klein and Lyytinen (1995). Their historical account focuses on the design process itself, while this article will be concerned with the development towards user participation in design. What is interesting in the latter historical development is how systems designers and worker participation initiatives bumped into one another and ultimately came to align themselves into a more or less symbiotic relationship. As Hirschheim and colleagues make abundantly clear, the various approaches to systems design have very different sorts of metaphysical, epistemological and normative assumptions behind them and objectives ahead of them. Yet despite

their ideological differences, their practices and responses to problems bear many similarities worth reflecting on.

This article will be more concerned with the journey toward this convergence than with the final destination. I will simply use the idea of a convergence of approaches to guide the narrative of the history leading up to a body of participatory design methodologies that all seem to share a value in explicitly "representing" users in the design of information systems. The various methodologies these approaches have arrived at fall under the broad description of "participatory design" and share this description by virtue of the fact that they each seek to integrate the end-users of an information system into the process by which that system is designed. A comparison of methodologies will also invite contrasts: in who the "user" is, in what part of "design" the user becomes involved in, in what goal "participation" is hoped to achieve, and in what the crucial aspects of that participation are. The real claim of this article is that over time, the concepts which originally divided design ideologies have started blending together, that this blending has occurred at the interface between human values and technological development, and that engaging this interface as a social and engineering problem has resulted in several common sorts of difficulties regardless of ideological perspective.

We will begin the history of this convergence by looking at the approach taken from the perspective of technological rationalization,1 and which arrives at user participation as necessary for efficient design. We then turn to the socialist and humanist approaches, which arrive at user participation in design as necessary for collective security and individual autonomy. Next we briefly examine the rhetoric of corporate restructuring and how it builds on the concepts of worker empowerment, and utilizes the methods of user participation, to legitimate new political regimes within organizations. Then we consider one of the various brave new hybrids of participatory design which tested the boundaries of political acceptability for changes in engineering reorganization. And finally we compare the representational practices of participatory design to cultural anthropology and draw some conclusions about the lessons to be learned from participatory design in considering a critical theory of technology.

The historical presentation traces two major traditions which have only recently converged into the rather heterogeneous field of practices that constitute participatory design. The first tradition I will examine is the development of the user-involving design methods which originated in large US corporations producing office technologies, and take IBM as my principle example. IBM's design practices represent a continuous tradition and practical methodology which began systematically involving users in a methodology first developed by systems designers in 1977. Called Joint Application Design (JAD®),2 it was derived as an extension of an existing IBM design methodology, Business Systems Planning (Carmel, Whitaker & George, 1993, p. 41). I begin the historical narrative with IBM because it represents a fairly linear extension of older rationalist and functionalist design methodologies. While not often recognized as a major contributor to participatory design, Joint Application Design not only addresses the integration of users into systems design, it also provides insight into the corporate culture which would later adopt variants of the participatory design methods originating in Europe.

The other, and better recognized, tradition which contributed to the current field of participatory design had its roots in the post-war work of social scientists at the Tavistock Institute of Human Relations in London, but really began its historical development in 1960 with a series of four labor organization experiments called the Norwegian Industrial Democracy Project. That project led to two different research programs: one in Britain, the "socio-technical systems" approach; and one in Scandinavia, the "collective resources" approach. These strains subsequently grew back together in the early 1980s but only recently found a broad influence in North America.3 The literature commonly refers to this tradition as the "Scandinavian approach," or simply Participatory Design,4 although it consists of many diverse techniques and methods developed by British as well as Scandinavian researchers.

One interest in examining the independent development of methodologies from these two perspectives is that, despite their differences, the approaches ultimately converged on a set of shared concerns and very similar practices. From a closer practical perspective, each design tradition

recognized a set of problems surrounding the position of the "user" in systems design. Yet the articulation and resolution of these problems took very different turns and expressed different values. From a broader cultural and political vantage, the two traditions have very different origins and maintain very different values, and it is thus surprising that they should find as much common ground as they do. There are, of course, critical issues which arise as the two traditions grow closer together, and the details of their historical development become crucial to any critical understanding of it.

After recounting the history of European Participatory Design, I will return to the transformations which occurred in American business and which further altered system design methodologies in the late-1980s. These transformations included the widespread adoption of participatory design principles and practices by North American corporations. This occurred first through the singular, though high-profile, work at Xerox Corporation, which was strongly influenced by the European researchers in the early 1980s, but continued to spread, and much more quickly, in the late-1980s. This pattern of growth coincided with the more general movement of corporate restructuring. From among many similarly relevant projects during this later period, I examine a single project to illustrate the impact of the changing conceptions of corporate organization on design methodology.

The methodology, called Engineering Codevelopment (EC), evolved through an experimental project sponsored by the Commission on Preservation and Access (a private, non-profit organization) begun in 1989. Called the "Class project," it was a joint venture between Xerox and the Cornell University libraries to develop proprietary digital-image technologies for the preservation of, and on-line access to, delicate rare books in the libraries' collections. Unlike the two principal traditions we will investigate, Joint Application Design and Participatory Design, it is not really a tradition so much as an exemplar of an interesting design perspective which lies somewhere in the field of convergence between these traditions. Xerox's design methodologies have been numerous and varied, arising out of more academically and experimentally oriented research than was typical of most American companies, focusing in the early 1980s on cognitive ergonomics and human–computer interaction at its Palo Alto Research Center (PARC), and by the late-1980s on systems design organization with its newly created Work Practice and Codevelopment Group. I examine only one of these methodologies in detail because it represents an interesting point of intersection between the Joint Application Design and European Participatory Design traditions—researchers there were influenced by both approaches. It thus serves as a useful illustration in understanding how participatory design reshapes design practices and, because it represents many of the features sought by management theorists in their discourse on Business Process Reengineering, it also provides an illustration of one way in which participatory design methods were able to align with this field of discourse. I choose to examine Xerox's Class project, despite Xerox's consideration of this project as a failure, because it provides insights into the complex social factors involved in technological design, and the ways in which these factors influence design practice.

The two main methodological approaches of Joint Application Design and European Participatory Design began with very different perspectives on technology and the role of technology in the workplace. This led to very different ways of conceiving of the "problem" of integrating technologies into the workplace. Moreover, each perspective and set of problems developed in very different contexts; these arose in different kinds of organizations—public, trade-union, commercial—and in situations which placed designers and users in very different relationships. There are two interesting phenomena which resulted from this and which I wish to emphasize. The first is that at some point, each of these traditions decided that systematically involving users in the design of technological systems was central to achieving their objectives. The second is that despite seeing the users' participation in the design process as an essential local objective, how each tradition conceived of the users' role in that process was shaped by their global objectives. This configuration turns out to bear a striking resemblance to that confronted by colonial anthropology, and its researchers came to address many of the same critical issues as those reflecting on participatory design. Finally, we will consider the role which technology should be recognized as playing in political discourse. With these issues in mind, I turn now to the background of these different approaches to user participation in systems development.

2. Technological rationalization: early 1970s to mid-1980s

2.1. Background

Before describing in detail the methods of participatory design, it may be helpful to review the basic methods of system design which these approaches sought to reform or replace. Arguably, information systems began with the first written records, probably Babylonian cuneiform impressions on clay tablets. But the profession of information systems design did not emerge until the 1950s, when computers first began to be applied to organizational information problems. Hirschheim et al. (1995) describe this early period as the "pre-methodological" era which was characterized by the "seat-of-the-pants approaches" to systems design (p. 29). There were no precedents to reflect on, and new methodologies grew in response to new challenges. Concerned primarily with programming and the management of physical data storage (e.g. which stack of punch-cards goes with this program), systems design was largely driven by technological considerations. From the beginning, large information systems projects were subjected to management techniques in a manner similar to other engineering organizations.5 Eventually, programmers, systems designers and management information systems analysts emerged as professional groups with recognized roles in systems development.

By the mid-1960s, some standardized methodologies had developed from these early approaches. They are best characterized as the "life-cycle" methodologies. The basic idea of these methodologies is that systems development ought to consist of a series of stages, which begins with defining a system's requirements, progresses to defining the data structures and algorithms necessary to realize the requirements, and then manages the actual programming and testing of the system. If the system's future users needed to be consulted during this process, this was done unsystematically through informal interviews, a process called "requirements gathering," which was done only in the initial, pre-planning stage of design, or in an evaluation of the finished system after it had been implemented.

2.2. The user as functional input: joint application design

The development of Joint Application Design (JAD) is itself indicative of the difficulties encountered by rationalistic system design methods. The conventional wisdom of systems design was embodied in Business Systems Planning, IBM's development methodology during the 1970s. A classic problem for any form of centralized planning in a hierarchical organization, where those doing the planning are removed from the activities of those the plans are being made for, is poor communication. In IBM's case, systems designers were finding it difficult to formulate system requirements from their labs while users were being frustrated by systems that failed to suit the needs of their office. Informed by insights from group dynamics and social psychology, Joint Application Design was developed in 1977 by IBM employees Chuck Morris and Tony Crawford as an extension of the existing design methodology. The intended objective of the methodology was to reduce the time required for the System Development Life Cycle (SDLC) while simultaneously increasing quality and reducing overall costs. The impetus for this methodology was thus only minimally theoretical and overwhelmingly technical.

To achieve these goals, the methodology sought to integrate structured meetings with users into the SDLC; it is these meetings which stand as the essential defining characteristic of Joint Application Design. The meetings are arranged so as to occur several times during the earlier stages of the SDLC, where meetings in the earliest stages focus on high-level user concerns and objectives while meetings in the later phases of design demand increasingly detailed information from users, with the ultimate goal of creating a single Design Document. The Joint Application Design Document is intended to provide the system's requirements and consists of a list of user requirements approved by everyone attending the meetings, thus constituting both an object of group consensus and a technical resource for design.

A universal feature of Joint Application Design meetings is that they are highly structured by a concern for maintaining social control in situations which might otherwise call into question the relationships between experts, managers, and workers, or digress into an unproductive chaos:

The Joint Application Design methodology emphasizes structure and agenda. This is evident in the JAD literature that reads somewhat like cookbooks. Everything is explained in great detail: "to do"

lists are included, as are masters of useful forms. There are four necessary building blocks for a well-run JAD meeting:

1.Facilitation. A designated leader (or leaders) manages the meeting. Some JAD practitioners consider the meeting leader to be key to process success, even more so than the act of gathering the users in one place, the essence of JAD.

2.Agenda setting/structure. The meeting must have a plan of action.

3.Documentation. One or more designated scribes carefully document everything in the meeting. Various lists are rigorously maintained.

4.Group dynamics. Group dynamics techniques are used for inspiring creativity (e.g. brainstorming), resolving disagreements (e.g. airing facts, documenting them as "issues," taking notes), and handling speaking protocols (e.g. enforcing "only one conversation at a time").

(Carmel et al., 1993, p. 41)

In addition to the facilitator and scribe, the key individuals involved in these meetings are users and designers. It is important to note that the "users" in these meetings are supposed to consist of managers and veteran workers with detailed knowledge of the work process. The implication is that "user satisfaction" consists in not only satisfying the requirements of the work process but in satisfying those in charge of overseeing and managing those processes. As Carmel et al. (1993) report:

we have observed numerous North American JAD meetings in which operational employees are overlooked as participants. This results in a meeting room filled with middle managers and supervisors unable to specify details of day-to-day operations (e.g. what 17 fields are needed to fill out form A345). This organizational failure stems in part from an unjustified lack of confidence that "front-line" workers can meaningfully contribute to the design process. (p. 46)

This raises many issues regarding the implications of information technology in organizational control and workplace politics which will be addressed more carefully in the final section.

Besides limiting the voice of the worker as a "user" through the explicitly management-dominated organization of meetings, Joint Application Design also serves to protect and promote the authority of technical experts. Indeed, the ostensive objectives of the designers are embodied in the meetings' structured nature by its simultaneously satisfying two functions: (1) the extraction of knowledge, beliefs, impressions, and desires from users in a controlled fashion through designer-established agendas, and (2) the rationalization or "selling" of the system to users by design engineers. Function 2 is achieved in part through the use of elaborate visual aids which seek to enable design experts to, among other things, describe their system to users and justify technical constraints6 and in part because everyone in attendance at the Joint Application Design sessions is considered to have "signed off" on the Design Document those sessions produced. The literature seeking to improve on the Joint Application Design methodology generally focuses on slight alterations in the presentation aids, forms, or the overall organization of the meetings themselves. The dual function of the Joint Application Design meeting allows the technical experts to represent users' needs as objective data in the technical design phase by using the information contained in the Joint Application Design document. Thus the technical design process may largely retain its rationalistic procedures while the users' influence on design is conveniently reduced to a well-structured functional input to the design process, a process which always remains in the control of the expert designers. It is precisely this highly structured nature of the process which is touted by how-to books on Joint Application Design (Wood & Silver, 1989; August, 1991; Crawford, 1994). So while Joint Application Design does seek to integrate users into the design process, it is unwilling to call into question or transform the fundamental technical rationality, practices, and political organization of that process.

3. Technologies of socialism and humanism: mid-1960s to mid-1980s

3.1. Background

European Participatory Design has its roots in a very different socio-political sensibility. The Norwegian Industrial Democracy Program consisted of four experiments carried out by researchers from the Tavistock and the Norwegian Work Research Institute between 1964 and 1967 (Emery &

Thorsrud, 1976). These studies investigated how social groups formed around production technologies and sought to reform job distribution and wage systems for workers. After these four experiments, two research programs developed along different trajectories: Scandinavian researchers focused on union empowerment through “collective resources” and British researchers focused on autonomy in work group organization through “socio-technical systems design.” Each felt that they had chosen the most promising set of objectives for what they saw as being feasible for the democratic reform of workplace technology and hence each saw the objectives of the other as tangential to the central issue. The British saw the union-centered approach as only being viable in the political environment of Scandinavia and as failing to theorize the organization of labor on a fundamental level, while the Scandinavians saw group dynamics as being ineffectual because it failed to consider the predominating power struggles of class and capital. Both approaches, however, were motivated by a shared concern for workplace democracy and the humanization of work and both contributed to the broader Quality of Working Life movement then beginning to take shape.7

3.2. Empowering the working class: collective resources

The Scandinavian collective resource approach originated when the Norwegian Computer Centre (NCC) began working with the Norwegian Iron and Metal Workers Union in 1970 to educate union officials on how technology affects working conditions and might be made to serve union interests. The expressed goal was to assist unions in devising technological control activities and policies. The basic methodology was to set up union mechanisms to gather and analyze information about specific technologies and their effects on workers. The belief was that by doing this the unions could offset the employers' natural advantage in technological knowledge and make it possible for the unions to put technological issues on the bargaining table. Research revolved around an elaborate “negotiation model” which sought to depict the bureaucratic process of introducing new technologies on the shop-floor in a way which would allow trade unions to intervene in response to management's technological proposals (Ehn & Kyng, 1987, p. 42). The “collective” here indicates that the intention was the empowerment not of individuals in their workplace but of the trade union collective in bargaining situations, while “resources” indicates the value placed in information resource gathering on the part of trade unions. The initial projects did not seek to integrate the workers directly into technological design processes, nor did they recognize design processes as a particularly significant locus of interest. In fact, worker participation (in the process of information gathering) was seen as problematic in that (1) the workers involved might become experts and join management thereby threatening union solidarity, (2) such a process could give management undue access to “shop floor information,” (3) it could prevent effective trade union participation, or (4) that it could even become a managerial strategy for worker manipulation (Ehn & Kyng, 1987, p. 40). Here the concern was not in the democratization of the technological design process, but of the bureaucratic decision process through which a company would seek to introduce a new technology on to the shop floor.

By the nature of the Norwegian Work Environment Act of 1977, which provided participatory rights to all (not just unionized) workers, issues of workplace democracy were seen as requiring locally specific actions and solutions resulting in several highly specific local projects. And even though the law provided for individuals' rights in workplace co-determination, due to the political relationship which existed between managers, workers and unions, only union-initiated activity was seen as having a viable impact on workplace organization. This was largely a consequence of the theoretical framework which motivated the researchers, who saw themselves as trying to find viable alternatives to the Tayloristic rationalization of work. Inspired by Marxist critiques of technological rationalization from authors such as Braverman, Noble and Winner (see Braverman, 1974, Noble, 1977, Noble, 1979 and Winner, 1977), they believed that unions were the only viable point of resistance to the otherwise inevitable capitalist processes of deskilling and increasingly centralized control through the division of labor, even though they expressed some doubts:

However, there were also practical and empirical anomalies that could not be explained by these theses. Work was not deskilled in all cases. More collective forms of work organization than the Tayloristic were sometimes proposed by management. It happened that workers gained from the introduction of new technology, etc. But this does not mean that the Marxist approach to

understanding changes of the labor process in a capitalist economy has to be rejected. (Ehn & Kyng, 1987, p. 36)

Yet it was precisely this insistence on the Marxist critique which would motivate their rejection of other theoretical approaches until the continuance of their own work necessitated a recognition of the significance of the technological design process. It was the explicit theorization of design processes that the Scandinavian researchers were to later discover in “socio-technical system design” and which would mend the schism between Scandinavian and British researchers after a decade of independent work.

Early collective resource studies targeted the impact of new heavy manufacturing technologies, which in themselves were not particularly flexible, and other areas where the unions had found it difficult to translate workers' interests into negotiable demands. Researchers and union officials sought to analyze workers' feelings about existing technologies (through surveys, union meetings and other methods adapted from psychoanalysis), to require employers to disclose information about technological reorganization proposals (such as the technical specifications and organizational policies involved in introducing numerically controlled machines and computer-based planning systems into the production line), and to produce textbooks and form classes for the education of union officials and workers (Ehn & Kyng, 1987, p. 28). These methods met with limited success in achieving those goals which could be easily formulated as collective demands, such as requiring the retraining of workers displaced by a new technology, but it was difficult to make any gains on qualitative humanistic concerns through bargaining. The approach failed to spread as a general union movement because most local unions could not spare the necessary time and financial resources required to make it work.8

A shift in industrial focus and the conception of design came about with the “second generation” of the Scandinavian approach marked by the Swedish–Danish UTOPIA project in 1981, the first recognizable participatory design development project. Conceived in response to the discouraging results of the earlier trade union projects which had found that existing technologies limited the possibilities of workers to influence workplace organization, UTOPIA targeted technological development as a prospective site for user involvement and influence. According to the Marxist critiques, the technological dehumanization of work through deskilling, intensified division of labor, imposition of rigid and routinized practices, and the shifting of control toward the top of organizations was an inevitable result of the introduction of new technologies which necessarily served the interests of management and owners. Since the existing technologies were presumably all being developed to satisfy the interests of their purchasers, the business owners, and hence to increase productivity, control, and efficiency, the only effective means of empowering workers in competitive industrial markets would be the creation of alternative technologies designed around workers' interests. Thus, the researchers sought to realize the ideals of the Scandinavian Technology Agreements established in the late-1970s, which had empowered workers with control over workplace technology in writing but not in practice, by designing a technological system with their skills and interests in mind (Greenbaum & Kyng, 1991, p. 11).

In cooperation with the Nordic Graphic Workers Union, the UTOPIA (both an acronym and an ideal) project studied a group of newspaper typographers working without computer support in order to develop a state-of-the-art graphics software product for skilled graphics workers. The objective was to create a commercial product which the unions could then demand as an alternative to other deskilling technologies. The commercial product was ultimately unsuccessful due to a small market and the shortcomings of the company which owned and marketed it. The completion of this project in 1985 coincided with a renewed acceptance of the socio-technical systems design research being conducted in Britain.

3.3. Autonomy of the work group: socio-technical systems

The British researchers were interested in the phenomena of group dynamics that had been observed in the early Tavistock inquiries into “leaderless groups,” originally motivated by founding member Eric Trist's personal fascination with the efficiency of German panzer tank divisions (Mumford, 1987, p. 61) and Britain's War Office Selection Boards' interest in choosing and training

military officers in order to utilize the phenomena of emergent group dynamics (Trist, 1993, p. 42; see also Bion, 1946). The actual experiments involved observing coal miners and how they organized their labor practices around technological systems. Miners working in large shafts were organized around assembly-line type machinery and showed little variation in their repetitive tasks, while miners working in small shafts or exposed faces, where the machinery could not be installed, developed novel dynamic and efficient work practices. Amenable to psychological and biological theory,9 the researchers developed their "organic" theoretical approach from systems theory, and focused particularly on the notion of "open systems," which itself drew heavily on a bio-organic metaphor (Mumford, 1987, p. 65). Thus work organization was theorized as an organic relationship between workers and technology—the socio-technical system—which ought to be analyzed according to criterion of "health" instead of the raw productive output measures of purely mechanistic analyses. From their studies of miners working under Tayloristic methods, they concluded that inefficiency resulted from optimizing the technical components of the system at the expense of the human components. The optimization of the socio-technical system as a whole would thus require a joint optimization of both aspects with an eye towards the social and psychological impact of technology on workers (Mumford, 1987, p. 63). Research focused on the concept of the "autonomous work group" in which workers were allowed to spontaneously develop their own work routines, make decisions, and change tasks with little or no supervision.10 It was this theoretical conception of the labor process that would motivate much of the later development of participatory design.

During the time that the Scandinavian researchers were concerned with empowering unions, the British researchers had been developing "socio-technical design principles" and management philosophies.11 It was just these projects which were at the time being criticized by Marxist theorists in Scandinavia for promoting values that were fundamentally capitalist—increasing productivity and decreasing worker resistance. While these criticisms carry some force with respect to how the principles were often applied within the workplace, they still depended on a mechanistic conception of labor process organization which pits capitalist demands against humanist concerns and leaves no real space for compromise.

It was against this same background of concerns that a broad international social movement, called the Quality of Working Life movement, began demanding more humane work environments and transforming the conception of workers' relation to their work away from a purely economic conception to one which included an emotional investment in, and personal attachment to, work. This was in part achieved through a reconception of the identity of the worker as an enterprising agent seeking personal fulfillment through job satisfaction. Miller and Rose (1995) have argued that by conceiving of the labor process as an organic process in which the most productive system was the healthiest and happiest system, socio-technical systems research made a significant contribution to this movement by offering a space for the negotiation of humanistic and economic concerns. They go on to argue that it is not as important to note the success or failure of the movement as to recognize the alignment of ethical, political, economic and technological elements through which it was possible for the Quality of Working Life movement to establish a particular identity for the worker and to restore the legitimacy of the corporation in industrial democracies.

The alignment of heterogeneous conceptions of work in the Quality of Working Life movement also opened a novel space for theorizing technology in which it could now be seen as serving multiple interests and values. So while it had been observed by Marxist critics like Noble and Braverman that technology could exploit and subjugate workers, it was now seen as also being able to promote workers' interests by making work more interesting, reintroducing skill, and by making practically feasible the "autonomous work group." The key to this lay in granting the worker direct control over the nature of the technology encountered in their day-to-day job. This also allowed the Scandinavian researchers to explain why the trade union approaches had failed—because the available technologies were not being designed in step with the new conception of workplace autonomy, union-supported technology had been just as mechanistic as management-supported technology. By 1985 the British and Scandinavian traditions had rejoined under a common banner of democratizing technological

systems design. The consequence was to be an increased emphasis on the involvement of the worker in technological design, which had already begun in the UTOPIA project. This was to be the essential feature of the tradition from that point on.

3.4. Participatory Design

Participatory Design researchers encountered two main barriers to the successful participation of users/workers in the design process. The first of these was a lack of appreciation by workers for their own knowledge of what they do—as one researcher reported:

It is a widespread opinion among workers that they themselves know nothing about technology, and that the necessary information must be obtained from management. This paralyzes the workers as far as actions are concerned. . . . [It] is at least as important to collect and prepare the knowledge of the workers, a knowledge they have obtained through their jobs. (Kensing quoted in Clement & Van den Besselaar, 1993, p. 29)

There was also a reluctance among technical experts to give project control to users, as this threatened their technical authority and traditional work practices. As a worker in one project remarked to the researchers: “But you don't always listen to us—you do what you think is right for us and the project. And, you are the expert; so who are we to dispute your decisions?” (Nurminen & Weir, 1991, p. 297). These two barriers together constituted the socio-political resistance to the democratization of the design process. Both users/workers and design experts found it difficult to leave their traditional socio-political roles in order to participate as intelligent and capable equals. The common response to this problem was to send the experts “into the field.” Rather than trying to rely exclusively upon special interviews or meetings to learn about users' work practices (as the union projects had done), researchers utilized “action research” methods whereby they spent a great deal of time observing and interacting with workers in their workplace. The reported consequences of this were an enhanced appreciation on the part of both workers and experts for workers' knowledge, and an increased understanding by workers of technology and its influence on their work practices.

Because the second-generation European Participatory Design projects sought to establish democratic participation among the workers influenced by a given technology, they saw their initial objective as breaking down traditional concepts of work and expertise among the design group (workers and experts). A second key concern for practitioners of European Participatory Design was enlisting external support for their projects and methods. Almost all projects reported resistance within or friction between the different organizations involved in the projects. Also, due to the nature of the legal, political and economic conditions (they were primarily academic and government initiatives) in which the European projects were situated, they sought to realize their objectives at a highly local contextual level, resulting in few immediately generalizable design principles. The confluence of these factors had led to a very nebulous concept of just what was entailed in utilizing a European Participatory Design methodology—it started to become an ideological approach rather than a prescriptive set of techniques. As a consequence of this, North American systems designers were very skeptical of European Participatory Design and worried that it was only viable in a context such as that which existed in Scandinavia and Britain. This is to say that as a design philosophy, during most of the 1980s European Participatory Design was unable to “sell itself” to a North American market still committed to its own tradition of software engineering practices and ideals. Skeptical system designers asked what methods it prescribed and what techniques it utilized, while European Participatory Design's proponents insisted that the key factors were the promotion of democratic ideals and an enormous creative effort on the part of designers.

4. Redesigning design: mid-1980s to mid-1990s

4.1. Background

Several changes in the late-1980s and early 1990s were to alter the way in which participatory design was perceived and which would both transform it and carry it across the Atlantic. Early European Participatory Design focused purely on democratic participation and overcoming various difficulties in achieving this; moreover, it was primarily conceived and pursued by an academic

community that held little concern for business interests and only limited responsibility for producing actual systems. This was in turn a very different set of problems than those which gave rise to Joint Application Design and similar methodologies in North American corporations, where the problems addressed by system design professionals focused on promoting business goals by increasing the efficiency and effectiveness of technical design. The consequence was that for more than a decade both Joint Application Design and European Participatory Design developed in isolation with almost no interaction between designers working in the two traditions. The interesting exception to this was Xerox PARC which was an early supporter of European Participatory Design in North America, though it stood in relative isolation in this regard. Joint Application Design grew up in the halls of North American industry12 and received very little academic attention, while European Participatory Design was conceived by academics and grew up in the progressive social democracies of Scandinavia.

In recent years, the two traditions have grown close enough together to find themselves publishing in the same professional journals and exchanging techniques and tools. This convergence was not the result of deep theoretical insights stemming from either side, but of separate movements by each toward filling a space opened by a newly emerging organizational regime. This new regime, described by Agre (1995) as the “empowerment and measurement regime,”13 was constituted by elements taken from management theories concerned with product quality and business process efficiency, accountancy methods seeking more precise cost and expenditure measurements, and nationalist political rhetoric over the global competitiveness and security of national industries and their workers. This regime not only transformed the discourse of management theory but, when combined with global information networks, it also made possible global factories and very large distributed control systems through modularized organization, outsourcing, and the electronic transmission of office communications, design requirements, software products, and even programming labor. Co-evolving with and informed by these changing conceptions of workers and business organization, the current heterogeneous field of participatory design claims the twin goals of increasing efficiency (of both technical experts and users) and increasing democracy (primarily for users).

4.2. Reinventing the corporation, representing the customer, redefining the worker

In the mid-1980s, North American business experienced a trend toward maximizing the efficiency and flexibility of its organization.14 This movement toward “reinventing the corporation”15 led to a series of hot topics in professional management literature, the most significant to participatory design being Total Quality Management (TQM) and Business Process Reengineering (BPR). While these management theories each claim principled distinctions between them, their historical development reflects their joint participation in a larger trend and they have been combined along with other elements under the broad rubric of corporate restructuring. More importantly, they each constitute a field of discourse which has given shape to the empowerment and measurement regime (Agre, 1995) and the impact of each can be seen distinctly on the face of participatory design literature.

Total Quality Management is a management approach conceived largely as a response to the perceived shortcomings of American productivity in comparison to Japanese industry in the 1980s, and is comprised of an enormous and heterogeneous body of literature. A recent book16 on Total Quality Management lists eleven management theories which have made significant contributions to Total Quality Management: Scientific Management, Group Dynamics, Training and Development, Achievement Motivation Theory, Employee Involvement, Linking-Pin Organizations, Socio-technical Systems, Organizational Development, Corporate Culture, New Leadership Theory, and Strategic Planning (Schmidt & Finnigan, 1992, pp. 13–23). This broad field, containing several familiar theories, is brought together in Total Quality Management as a business and management philosophy that promotes the conception of a “customer-centered” business organization which is subjected to intense scrutiny and measurement. The objectives are to instill a conception of customers and their needs at every organizational level and production process in the company. This includes disseminating information about customer needs and the company's market position and objectives to workers at every level of the company, as well as benchmarking individual processes and products

against those of like competitors. It is the provision of benchmarking performance and quality reports to teams and individuals which is supposed to give them a tangible sense of their participation in the company's objectives, i.e. the satisfaction of the customer, and to allow them to set their own performance standards and reach their greatest productive potentials. Total Quality Management thus epitomizes Agre's (1995) conception of the empowerment and measurement regime by giving workers a sense of accomplishment in their work while subjecting them to increasing degrees of scrutiny and productivity demands.

Business Process Reengineering seeks a radical reorganization of offices, departments, and entire companies around specific “business processes.” One method employed is task-analysis, which seeks to establish the “process flow” of business functions, analyze out the distinct tasks and their functional relations, and logically reintegrate these into a more nimble and efficient business machine (Sachs, 1995, pp. 38–39). The reengineered processes are argued to provide workers with a more direct and enterprising relationship with their work and a stronger identification with their corporate culture. Much like Total Quality Management, Business Process Reengineering seeks to reorganize processes around the customer, and does so by creating entrepreneurial teams which are treated as autonomous and accountable entities within the company. The principal difference between the two is that Business Process Reengineering calls for a radical transformation of work organization and the subsequent elimination of jobs that this necessitates (Willmott & Wray-Bliss, 1995).

The applications of Total Quality Management and Business Process Reengineering in American companies have met with mixed success generally, but seem to have been particularly successful in establishing customer-centered product development processes in office and information technology industries. While the reengineering process has been realized in different ways in different companies, a Digital Electronic Corporation design-leader's experience is not unusual:

In the summer of 1993, the central engineering organization in Digital began the implementation of a re-engineering effort under the name of Achieving Engineering Excellence (AEE). This in turn was an integral element of Digital's overall re-engineering effort, inspired by a desire to streamline all our dealings with customers. A major goal of this effort was the reduction by 50% in new product development cycles. Internal data collected within Digital showed that the most significant contributor to excessive development cycles was a phenomena known as “requirements churn.”

The AEE data, gleaned from a survey of hundreds of Digital's staff and an analysis of the corporate planning database, found that on average, 40% of the requirements specified in the Feasibility and Requirements Phase of the Lifecycle were redefined in the subsequent four Lifecycle Phases. The cost of requirements churn, using an industry-wide regression model, found that on average Digital spent 50% more than budgeted. (Hutchings & Knox, 1995, pp. 72–73)

This led to the development of a Requirements Management reengineering team which sought to eliminate the “churn” by forming cross-functional design groups (with members from marketing, service, management, and customer representatives in addition to engineers), restructuring the requirements-gathering process into an iterative “listen–define–validate” model that relies on continual feedback instead of an initial exhaustive establishment of requirements (Hutchings & Knox, 1995, p. 74). Though not all companies approached this new problem in the same way, this is essentially the basic process by which many North American companies began using more participatory design methodologies.17 Through a demand for the representation of the user/customer in the design process, Business Process Reengineering and Total Quality Management transformed both the problematic of design and the standard of value for judging the design process and its products.

4.3. The autonomous engineer: Engineering Codevelopment

In many ways, Xerox's Engineering Codevelopment (EC) exemplifies the impact of socio-technical system design and customer-centeredness on North American systems design. Various other projects could serve this purpose as well—see, for example, Holtzblatt & Beyer (1995)—but Engineering Codevelopment does a good job of bringing together many of the themes and issues with which this paper is concerned. Engineering Codevelopment was more daring than most design methodologies in that it recognized that the successful integration of the user/customer in

that process will necessarily transform the work practices of the engineers. It approached the methodological problems of design with essentially the same general objectives sought by reengineering—reducing development time, improving product quality and customer satisfaction—but with a slightly different set of values. Besides improving the product through the integration of users in design, Engineering Codevelopment sought to improve the practices and skills of engineers by engaging them in novel situations where traditional practices and routines cannot be readily applied:

Working directly with users and supporting their day-to-day work require engineers to be committed to helping users in a personal way. Hierarchical dependency relationships between engineers and managers do not work in a codevelopment effort that bridges two different enterprises. In addition to independence, team members are encouraged to develop a diverse set of technical, interpersonal, and often interdisciplinary skills. Finally, individuals develop their own direct and informal contacts within their own and the users' organizations. (Anderson & Crocca, 1993, p. 49)

This “radicalization” of engineers' work was not in fact an objective at the beginning of the experimental project, and was one of the causes of the project's failure. Thus it was neither arbitrary nor purely philanthropic, but reflected Xerox's committed objective to experiment with and develop improved engineering methodologies—they were willing to try it in order to “see what would happen.” Xerox's desire to participate in the Class project with Cornell librarians was primarily motivated by a theoretical rather than a purely practical concern, but one which still sought to achieve business objectives valued by an office technologies industry seeking to improve its technological advantage through organizational change. It should also be noted that the Class project was undertaken by Xerox researchers outside of PARC, though PARC researchers did consult on the project.

The Engineering Codevelopment methodology was characterized by two essential features. The first of these was a customer-centered prototyping methodology. The system being designed was a scanning and retrieval system for some 1000 brittle books in Cornell's libraries. The project consisted not just in designing an information archiving system, but in the development of the physical digital scanning and printing devices necessary for its effective use. Hence, the methodology placed a great deal of emphasis on obtaining customer reactions to working prototypes placed in the customers' workplace. The idea was thus to “tune” the artifact to the work environment in which it was to serve. In addition, Engineering Codevelopment was characterized by its willingness to let engineering practices develop open-endedly around the requirements of prototype development. This was reflected in the enormous amount of autonomy granted to the design team by their own corporate management. Where traditional engineering methods are organized according to a hierarchical and functional structure, the division of design labor tends to be rigidly controlled by management, who maintain authority in key decision-making situations. The semi-autonomous Engineering Codevelopment group dissolved its internal organizational hierarchy and also obtained some degree of executive autonomy from their corporate supervisors (Anderson & Crocca, 1993, p. 54). This limited autonomy resulted in a profound transformation of engineering practices in the design process, and it also contributed to the political tensions which ultimately led to the project's poor reception within the company.

Traditional design methods at Xerox place marketing research and development groups in charge of producing detailed system specifications, often through survey research. These specifications are in turn transmitted from marketing to design groups and subgroups via the established management hierarchy and functional sub-divisions in the engineering organization. Engineering Codevelopment, by contrast, seeks to develop a system for a customer with needs that cannot be formalized by market research methods. The design team is thus charged with the novel task of evoking and verifying the system requirements through personal interactions with users. The idea is to establish the needs of users and codevelop the prototype in real time by letting users judge the adequacy of the prototype's features (rather than test engineers as is traditional).

The consequence of this new demand on the designers in such a situation, at least in the Class project, was a radical transformation of engineering practices. Whereas design traditionally begins

with the division of tasks and responsibilities among the design team, the Engineering Codevelopment group members "float" among tasks and leadership roles through the course of the project based on personal initiative and experience. This had five advantages as reported by group members: (1) decisions were made by those most involved with a problem rather than by a manager with no antecedent knowledge of the matter; (2) because designers are not locked inside functional "black boxes" they are each aware of the overall situation at hand and how their activities are situated within it, resulting in a better overall design; (3) designers were able to participate in diverse tasks which broadened their skills and knowledge while enriching their work experience; (4) all members are fully employed and do not spend large amounts of time waiting for others to complete their tasks before they can begin; and (5) group members are chosen and valued for their particular and unique skills rather than their conformity to a homogenizing structure (Anderson & Crocca, 1993, p. 54).

In contrast to Joint Application Design methodologies which sought to alter the hierarchical and functional design process by merely adding a new functional component to that process, Xerox's Engineering Codevelopment was willing to forego the organizational structure of its design process in the hope that the flexibility provided would produce a working system where its traditional market research methods could not reach. While Xerox's explicit objectives were to design a highly customized system for their customer, they were also interested in studying the novel methods and group dynamics that would evolve in a design team organized around a customer-centered design process and granted a great deal of autonomy. The experiment not only resulted in a product which satisfied the customer but also in a reconstitution of the engineers' work organization, which led to the promotion of some empowering humanistic values—job satisfaction, diversity of experience, skill appreciation, personal autonomy, and educational development.

It is worthwhile to note that the project's leader, Bill Anderson, began his career at the Tavistock, and that the account of engineering practice presented as an outcome of this project echoes the virtues of the "autonomous workgroup." More importantly, the promotion of these humanistic values was limited to the engineers of the technological system, and did not extend to the users/customers. While interactions with the users stimulated this organizational transformation, the users' organizational situation was not profoundly changed. The introduction of a new technological system in the librarians' workplace certainly transformed their work practices in various ways, but from a design perspective there was an established objective to maintain a value-neutral approach to these transformations: "Even though engineers are changing the customer's work practice, they need to avoid interfering with the social and political dynamics that characterize that workplace" (Anderson & Crocca, 1993, p. 55). The result of pursuing such a methodology was a work practice very much like the ideal "autonomous workgroup" originally sought by the socio-technical systems researchers at the Tavistock, yet here it has again shifted its focus toward the design process and away from the consequences of technological change on the reorganization of worker's practices. And, like the European Participatory Design researchers, Engineering Codevelopment recognized the difficulties of communicating design concepts between engineers and librarians.

5. Reflections on technology and critical theory

Participatory design emerged at the convergence of two approaches: (1) a critical project which sought to rectify political imbalances caused by technologies in the workplace and to protect workers from technological change, and (2) the evolution of a technological rationalism which sought to increase the success and efficiency of new systems. We have just recounted the historical development of that convergence and noted the moments at which its purposes and practices shifted to accommodate new concerns and to abandon futile efforts. What can this history of participatory design tell us about the social implications of technology, and the possible roles for a critical theory of technology?

There are a great many issues bundled up in this history, and many of the individual strains have been discussed in isolation by various authors. I find such isolated discussions dissatisfying because, to borrow Heidegger's words, they leave us "utterly blind to the essence of technology." These discussions take on several forms in which technology is either a political tool or politically neutral, and if a political tool it is either completely plastic and subject to the predominating political will

(whether this be derived from underlying political structures or a consequence of political negotiation) or it is a rigid and unbending servant of political hegemony. None of these is quite right, and teasing apart the real essence of technology must confront both its empirical and political aspects.

In this final section, I will begin by examining the nature of "representation" as it plays out in the "representation of users" through the history of participatory design. This constitutes the scientific or empirical side of participatory design—its material and practical consequences for systems design. I will then turn to a discussion of the role of technology in political theories, and in particular the perspectives on technology taken by critical theory.

The conclusion I will reach is that representation is neither purely objective nor subjective, but is ultimately pragmatic. And technology is neither politically neutral nor deterministic, nor is it a perfectly plastic media waiting to be molded by political forces, but different technologies are more or less plastic and subject to being inscribed with political ideologies or enforcing political policies. Accordingly, technological artifacts are able to stand as shared referents by virtue of their material and practical consequences, and when engaged in a dialectic they offer resistances of their own which must be dealt with. A given technology will only be empirically and politically successful if it is able to survive a dialectic of design and use. While it is possible to get a technology "right" the first time around, the best guarantee of a technology's success is to subject it to successive redesigns informed by user reactions. The advantage of this dialectic approach is that it is able to address empirical and political, material and symbolic, issues simultaneously at each iteration. A system will be a failure if it cannot achieve the intended design goals, if it is unreliable or breaks internally, if it never gets used as intended or at all, or if it actually impedes the jobs of workers. Thus, participatory design methods can be a highly successful way to build technological systems because it integrates an assessment of material, practical and political consequences of a system in a single dialectic of resistance and accommodation.

5.1. Representing users: the science of participatory design

European Participatory Design had always been concerned with how technology alters work practices. At its inception, however, it did not consider the technological design process itself to be a key point of interest. Once projects like UTOPIA had begun to problematize technological design, they saw their challenges as being the overcoming of traditional roles, power relations and preconceptions of designers and users. Very quickly they added to this the problem of "communication." What researchers found most difficult, once socio-political barriers had been bridged, was that designers and users tended to talk past one another. Similarly, researchers taking the approach of technological rationalism to design thought initially that knowledge of users' requirements could be acquired straightforwardly through organized meetings. They learned quickly that this process was not so straightforward, and the make-up of a Joint Application Design meeting reflects a recognition of the difficulty of achieving this understanding—the need for visual aids, the importance of "scribes" to record "lists" of priorities and requirements which arise during meetings. These approaches thus shared a recognition of the problem of communicating technical designs early on.

Historically, representing the practices and needs of workers grew out of the old and adversarial traditions of worker advocacy by unions, and the Tayloristic analysis of work practices. Producing such representations has been a long-standing and controversial enterprise. On the one hand, expertise and skill are identified as the personal traits of skilled individuals, and are therefore are a human value which is alienated from the individual in the process of extracting that knowledge and producing a representation of it. On the other hand, there are the needs of workers which the union attempts to represent as a political agent acting on the behalf of the worker. In the case of skilled workers in a factory, these needs and expertise can often be clearly delineated. In the case of information-intensive office work, this task can be more difficult: when is certain information required by a certain task and when is access, or the privacy from others having access, to that information a right of the worker? When is a task a valuable skill, and when is it merely a burdensome chore? The rationalization of work and the political needs of workers are themselves difficult to

represent, and in the process of information systems design the problems frequently seem to mix together and become even more difficult to get a grip on.

A great deal of literature has been produced by both approaches on the problem of communicating the needs of users. In fact, this problem can be seen as an aspect of the problem of representing users' skilled practices and needs in general, where asking them to communicate this knowledge is simply the approach taken by participatory design. For example, the essence of Joint Application Design is to involve users in meetings during the requirements gathering phase of the system development project—the desired objects of knowledge are the requirements, and the users are a means to this end.

Researchers also encountered problems in uncovering knowledge of tacit skills or embodied routines, and realized a necessity for respecting the fact that workplaces have a rich local vocabulary that takes time to master and is not always easily translated to individuals outside the workplace. Thus there emerged a recognition of the problem of communicating users' knowledge to engineers. It was expected that bringing the users into the design process would also bring their tacit knowledge into the technological product. This turned out to be more difficult than expected, and the different approaches responded differently to this challenge. The technological rationalists saw the problem as one of properly "representing the user"— that there was some objective knowledge held by the user which needed to be elicited, but there was resistance to this elicitation, either in the users or in the system of communication between users and engineers.

From the socialist and humanist perspectives, there did not exist the same faith in the existence of some objective knowledge on the part of users about their own skills and practices, but there was a recognition of the politically charged nature of the workplace and any interactions which might occur between engineers and users. They believed that some users may not feel politically safe in articulating their needs, or that the needs of a collective of workers were essentially a coalition of different and sometimes conflicting interests which had to be negotiated among participants. The solution in both cases was to bring users and engineers into closer practical interactions—in structured and unstructured meetings, and by sending engineers into the workplace to observe users or even participate in the work practices of users, as users.18

The problem of representing the worker, and its ambivalent status, has been well articulated by Suchman (1995). Suchman offers several key points for reflection to social scientists and systems designers involved in representing workers for the support of work. These points stem from the recognition that attempts at empirical objectivity and accuracy often neglect political realities. One point is that the ways in which representations are utilized in design can limit user autonomy regardless of their accuracy. The principal limitation on autonomy presented by information systems is the managerial control and surveillance that is made possible by these systems. Once a task is embodied within an information system, the observation of that task or of its products, and thus their measurement, becomes an easy prospect for managers and administrators.19 Thus, system designers need to be aware of the potential applications of their work even if they themselves have no ill intentions. It may also be the case that by the embodiment of a task in a technological system one limits the personal autonomy of the worker. However flexible they are intended to be, a computer system always offers only a limited set of ways to do things and by requiring workers to proceed in certain specific ways it can limit their ability to organize their other tasks, thus limiting their autonomous space of task scheduling. In short, regardless of the political or other intentions behind designs based on knowledge obtained from workers, the use of that knowledge has practical consequences and these will not necessarily conform to the design intentions, whether practical or pragmatic.

The notion of "representing the user" is further complicated by participatory design insofar as workers are expected to "speak for themselves." The next section will make a more careful consideration of the nature of user participation as one of political participation, but we should note here that the very idea of "representing the user" can be seen as a way of silencing the user. Moreover, not all of a system's users will ever be present during the design process. Only some will be present, and new users will have to be trained as they enter the workplace. Thus, even when users are fully involved in design, they are only representatives of a larger group of potential users. Still, there are

challenges to the full participation of these user representatives, not the least of which is conveying to them the significance of various design alternatives.

The nature of the problem of communicating designs to users and its solutions revolve around the different practical requirements for a design. What researchers found to be most difficult in communicating a technical system's design to users was that users lacked technical knowledge, and system designs are typically expressed in highly technical form. A formalized design made by and for engineers will amount to instructions and requirements for the necessary components and their functional interactions with one another—a technical "blueprint." An engineer's design specification will rarely describe features in terms of a user's actions, which are assumed to be given or implied by a "good" design—the blueprint of a house rarely depicts its occupants. To a user, the internal functional specification is almost meaningless, as their concern is with the practical activities it supports and the ways in which their own practices will be altered by a new system. Since users were not participants in the actual technical execution of the design, they found it difficult to understand the various system designs which the engineers proposed. What was needed was a new way to represent a system design which did not require technical knowledge to interpret its practical consequences. The various prototyping and visualization methods developed by participatory design researchers thus attempt to build a bridge from the engineers' design alternatives to the users needs by creating an intermediate representation which is technically feasible and affords practical interpretation.

The initial attempt at this kind of representation in the UTOPIA project was to develop exemplary "screen shots" of what the potential design would look like. This helped, but users found it impossible to judge whether such a design would satisfy their needs since they could not conceive of how it would actually operate. In response to this problem, the researchers developed many innovative means for communicating the practical functionality of various designs and design elements to potential users. The participatory design literature has since produced numerous articles on "prototyping," "visualization," "mock-ups," "storyboarding," "metaphorical design," and "future workshops" which all have the expressed purpose of offering suggestions of how to develop and use videos, transparencies, functional prototypes, and even cardboard boxes and plywood to give users a sense of how a proposed system will work.

It is interesting to note the similarity of many of the early European Participatory Design visualization techniques and future workshops to the techniques and meetings developed by Joint Application Design practitioners over a decade earlier, though there appears to have been no direct influence from the older Joint Application Design tradition on the European researchers. Xerox's Engineering Codevelopment takes the methods of technological imagination one step further by introducing the working prototypes into the user environment. This final step, however, turns out to be qualitatively different from the others insofar as users interact directly with a prototype of the new technology, rather than with a representation which is still subject to interpretation by an engineer.

When taken together, we can see in these two "problems of communication" an image of the basic mode of scientific knowledge. The image that science extracts objective knowledge from the world is highly problematic, but a discussion of this metaphysical issue is beyond the scope of this article (see Latour & Woolgar, 1979; Hacking, 1983; Pickering, 1995). Let it suffice to say that according to this image, there is a matter of fact out in the world which is observed, understood and assimilated into a theory, the theory has certain consequences, and these consequences can be reliably expected to occur under specific circumstances. In the way that a doctor discusses symptoms with a patient and prescribes a remedy, or a physicist probes a system and predicts the outcomes of future manipulations, the system designers ought to elicit knowledge of workers and prescribe a better system. There is, of course, a great deal of noise in both directions of this communication model; understanding a patient's symptoms, and getting a patient to take a prescribed medicine, setting up and interpreting instrumental measurements, and getting a system to react properly to an intervention, eliciting system requirements from workers, and getting workers to use a new system proficiently are all difficult tasks. But this model of scientific understanding and expertise is only an ideal, and there is really much more complexity to the situation. Indeed, it was been well argued that much of the work of

science involves making what actually happens in the laboratory "fit" with this ideal of what ought to happen (Pickering, 1995).

Many would argue that the comparison to natural sciences is unfair, and that systems design is really more of a human science in which knowledge is inextricably subjective. Indeed, the objective model of science is itself only a caricature which fails to take account of the many social and material negotiations which take place in the process of producing and exercising scientific knowledge, while social sciences are concerned primarily with these negotiations. The epistemic problem is not one of correctly communicating information from object of inquiry to the understanding subject, but as one of bridge-building—synthesizing a new field of symbolic meaning through a series of symbolic and material interactions. In the case of systems design, the interactive process is between engineers and users, and knowledge of system requirements is the outcome of the confluence of two different fields of symbolic discourse and sets of material practices. The objective of user participation in this situation is to align the concepts and representations of both workers and engineers around a common discourse and set of practices through which the desired technological artifact can emerge, evolve and become useful.

It is thus helpful to consider the case of a social science which has struggled a great deal with issues of representation—cultural anthropology and ethnography. Systems designers themselves have recognized the relevance of ethnographic techniques to their work. In fact, an entire sub-discipline of systems design called the "Ethnography of Information Systems" has emerged which identifies the role of the systems designer, or at least one role of a member of a design team, as being fundamentally that of an anthropologist—to produce a representation of the practices of a work culture which can be used as a basis for systems design. These representations are sought out as a basis for the reorganization of observed work practices in order to increase efficiency and productivity, as well as workers' job satisfaction.

One recent article from this discipline even echoes some themes from the early Tayloristic practices of filming workers to identify and isolate the key motions of their jobs. Produced by Xerox's Work Practice and Codevelopment Group, the article outlines and analyses the use of video for the recording and analyzing of office workers (Brun-Cottan & Wall, 1995). Instead of analyzing the kinesthetic properties of tasks as in Taylor's analyses, the videos are viewed and analyzed by ethnographers to reveal the hidden aspects of a culturally informed work practice.20 What is important to understand about this kind of representation is that it is necessarily reductive and for some purpose. A representation is valuable for what it leaves out as much as for what it contains, and the practice of creating representations does not seek completeness or objectivity but practical usefulness. It was precisely these kinds of practical representations which anthropologists came to recognize as the product of their profession. Many anthropologists felt confronted by an ethical crisis because such representations had been generated so readily to serve the interests of colonial control in many parts of the world.

Clifford (1986) has insightfully articulated the anxieties faced by cultural anthropologists reflecting on their role in the era of colonialism. His analysis focuses on the role of the ethnographer in representing another culture through interactions, photographs and writings in particular. The intriguing comparison to participatory design which strikes me in Clifford's account is that he conceives of the text as a consequence of a series of interactions—broadly construed as being between two distinct cultures, and narrowly construed as specific interactions between individuals. Under either construal, the text emerges as a synthesis of different perspectives which does not necessarily take on a single perspective. There are different ways to interpret the resulting text; in colonial anthropology the resulting texts can be criticized for their attempts to impose an imperialist perspective and take ownership of the knowledge of another culture by reconciling or exoticizing the differences between two cultures according to the purposes motivating authorship—the politics of colonialism. Similarly, the representations of workers which inform and justify the reorganization of work are criticized by authors such as Braverman (1974), Noble, 1977 and Noble, 1979 and Winner, 1977 and Winner, 1995 and Suchman (1995) as having the potential to disempower workers. Under such an interpretation, participatory design can be seen as a form of technological colonialism.

Another way to interpret the interaction between cultures is that a novel text is produced which does not have an identifiable perspective, but instead is decentered and multi-vocal. This is the way in which Clifford (1986) argues that ethnography can be redeemed from colonialism. By making explicit that different voices, experiences and perspectives are participating in the text, the reader is allowed access to the processes of interaction between cultures that the anthropologist has, without being required to draw the same conclusions.

In Joint Application Design, for instance, the Design Document could be interpreted as just such a multi-vocal text, representing not the objective needs of workers or system requirements but the negotiated outcome of interactions between users and system designers. Whether this is a legitimate interpretation depends very much on the context in which the document is produced and the practical consequences of systems which utilize that representation. The discourse between users and system designers is mediated and permeated by political imbalances, and lies somewhere on a continuum between the extremes of participants acting as free and capable equals in what Habermas (1990) calls an "ideal speech situation," and participants acting as interrogator and informant. As Hirschheim et al. (1995) make clear, the actual stance taken towards the explicit knowledge extracted through interactions between system designers and users is dependent on the paradigmatic assumptions of the system designers who use the knowledge to build systems. Whoever organizes the system development project and organizes the interactions between users and designers does so according to their own assumptions about what the goals of that interaction ought to be: technical requirements, user needs, political reconciliation, worker empowerment, etc. Of course, successfully imposing metaphysical and political assumptions on the design process depends on the success of attempts to maintain social roles within such meetings.

The notion of a technology being like a text at the point of intersection between two cultures has been well articulated in Star's (1989) concept of a "boundary object." But while information technologies do support the inscriptions and articulations of incongruent or even incommensurable perspectives and interests, Clifford's notion of the ethnographic text can add something more subtle to our understanding of participatory design. What he presents is a way to understand ethnography as a performance:

Cultures are not scientific "objects" (assuming such things exist, even in the natural sciences). Culture, and our views of "it," are produced historically, and are actively contested. There is no whole picture that can be "filled in," since the perception and filling of a gap leads to the awareness of other gaps. . . . If "culture" is not an object to be described, neither is it a unified corpus of symbols and meanings that can be definitively interpreted. Culture is contested, temporal, and emergent. Representation and explanation—both by insiders and outsiders—is implicated in this emergence. (Clifford, 1986, pp. 18–19)

Understood as a temporally emergent performance, ethnography becomes a way of interacting with the world centered around the production of a material record of those interactions. The same can be said regarding the processes of design which provide for interactions between users and engineers—the result in a system which stands as a material archive of their interactions with each other.

With the introduction of functional and working prototypes into the user environment, and thus into the interactions between system designers and users, the dialog between users and designers becomes dialectic. In cultural anthropology, the cultural insiders are rarely given the chance to challenge the representations made by the anthropologists' text. If the outcome of an interactive process of systems design is merely a document, it is still the prerogative of systems designers to produce whatever system they please, and it is the prerogative of users to resist whatever system the designers attempt to impose on their work practices. But when the technological artifact itself becomes part of the interaction, the material and practical consequences of design are reworked in the process. Designers are no longer attempting to interpret design requirements in isolation, but must respond directly to the users' reactions to the consequences of the existing system. Users cannot so easily reject a system as being remote from their needs and ineffectual when it has gone through several revisions motivated by their own challenges to its relevance and usefulness. And throughout

the redesign process, the practical and material resistances of the technology itself to being reworked in certain configurations becomes manifest to users and designers.

We have thus seen how two different perspectives on information systems design, one approaching from the side of the technological rationality and one from the side of social and human empowerment, have converged upon the same set of problems. By arriving at the point of contact between the human and the technological as the source of resistance in the development of information systems, both perspectives converge upon the problems of understanding and incorporating the practices and needs of the user in the process of developing new technologies. The consequence is that social progress and technological progress both come to socio-technological progress, and arrive at a common set of problems as a result.

5.2. Technology as politics: the critical theory of technology

In the end, participatory design went beyond the initial suggestions of written criticism to become an agent of technological change in the workplace. Many of its methods and ends were shared by the Quality of Working Life movement, and ultimately the corporate restructuring of the late-1980s. How are we to understand these critical and technological transformations and what can be learned from them? We could note the weakening of the critique itself—and we might try to explain this by appealing to the realities of business and technology as having confronted and overtaken political ideals. Or we could interpret the participatory and ethnographic approaches to systems design as enabling the regimes of empowerment and measurement and new forms of technological colonialism. Or we could instead note the success of enlightened design practice as one of the factors contributing to economic success and job satisfaction for many corporations and their employees. The development of participatory design is complicated by the historical turns it took and the diversity of projects that carry its banner, but it is hoped that careful attention to these complications provides an appreciation for the tensions we find in the current practice of participatory design. It is also hoped that the study of this particular historical strand of technological criticism can provide some lessons for technological criticism in general. What might this history tell us about the role for a critical theory of information technology?

The original motivations of researchers in the early Scandinavian projects was explicitly to counteract the dehumanizing effects of an increasing technological presence in the workplace. This was principally informed by a Marxist critique of the labor process and targeted Tayloristic rationalization as an assault on the human aspects of this process. Early Scandinavian researchers were also critical of the Tavistock researchers' desire to design and optimize the social side of socio-technical systems for being another form of the furthering of capitalist goals at the expense of workers—an objective toward making workers less resistant and more efficient. This disagreement over what counts as “democratizing work” exemplifies different conceptions of the nature of technology itself.

Feenberg (1991) is critical of traditional Marxist critiques for essentializing technological bias—the belief that technology has an essential bias towards serving the values of the capitalist society which produces it and thus that truly democratic technology can only be produced after revolutionary social reformation. He is equally critical of viewing technology as being essentially neutral and thus develops a notion of “ambivalent technology”, stressing its instrumental nature in support of any values that one wishes to build into it, which suggests that democratic social reformation can itself be affected by developing technologies which embody democratic values. This is essentially the same position that the Scandinavian researchers had finally arrived at with the second-generation projects and their renewed acceptance of the Tavistock researchers. It is important to note the way in which this view assumes that technologies are perfectly plastic and subject to human interests.

The shift is thus from a generalized resistance to new technology to the belief that some technologies may be genuinely progressive, and with this shift comes the insistence on introducing democratic ideals into technological design. But what are these ideals? Feenberg identifies essentially the same positive ideals that the Scandinavian researchers gleaned from Braverman's negative points: recontextualizing design, respect for the promotion and preservation of skill, the reintegration of

aesthetic and educational value in work, and anti-hierarchical and peer-oriented interaction. The ideals are certainly democratic, but in their abstract form how are they to be effectively realized in an actual technological system? The design process is an enormously complex one, and in any actual system there will be other values of utility and efficiency in play and possibly competing with democratic ones. In fact, it is precisely the issue of what counts as efficient and useful in a technology that problematizes requirements gathering and led even the most techno-rational corporate cultures to seek out the methods of social sciences for assistance.

It is these competing and overlapping values which meet in the current methods of participatory design and it is within specific design contexts that these values must be weighed out and matched to specific proposals and technological possibilities. Every design process is thus permeated with issues of value and how these will be decided is contextually dependent on work and process organization, power relations among individuals which implicate expertise, information access, authority, and rhetorical skill, restrictions on time and financial resources, and a myriad of incidental factors, not the least of which are the actual material capabilities and practical demands of the technology in question. Besides promoting abstract democratic values generally and in the design process, there would seem to be no specific, formal or structural prescriptions available for the "democratization of work" through design, and such a view ignores the possible limits of technological capabilities.

Despite these complexities in the design process, participatory design has been argued to stand as a model of how critical theory might approach technology. Some authors, such as Winner (1995), have proposed that participatory design might bring us one step closer to realizing a humanist ideal of Enlightenment democracy—a world in which people are universally empowered to determine the rules which govern their social practices—by providing a forum for bringing technological choice to the people. This was much of the motivation behind the Scandinavian participatory design researchers' interests in worker empowerment, but what does this assume about the political role of technology?

First, it is important to note that Winner (1995) is primarily concerned with the participation of users in discussions of system requirements, not necessarily with the accommodation of their reactions to working prototypes. Ultimately, politics are played out between the participants in the public discourse, and technology joins the politically marginalized as a pawn in the political game—having no agency as a political actor in its own right. Such a conception is difficult to maintain when we consider the technologies that Scandinavian researchers first attempted to subject to political control—heavy machinery—only to discover that it was not plastic enough to be the negotiated subject of collective bargaining.

It is also crucial to note that Winner is holding out a procedural notion of justice as his political ideal—it is the very participation of people in design that is democratic, just as the right of all citizens to vote makes a nation democratic. It does not follow necessarily from universal participation that the society or technology which results will be free or empowering—just as people are free to elect a tyrant or the right to vote may not come with other human rights attached. Instead, like "due process," participatory design is argued to provide a process which is the realization of justice even when the outcomes of the process may not be found agreeable to everyone. What Winner and many of the authors who promote Habermas' (1984) critical theory focus on is the centrality of public discourse to politics. This is to say that the rational political ideal is the participation of individual voices in debating and arriving at a mutual agreement on normative issues. The challenge facing such views is that participants in such discourse are never able to escape their power relationships with one another in order to participate on an equal footing in an "ideal speech situation" (Habermas, 1990) from which genuinely universal normative claims could be arrived at.

But I think such public discourse theories of politics suffer from another sort of idealism as well when they address technology issues. Consider a fundamental political issue such as the distribution of goods. Some of the most basic goods happen to be material—food, shelter, clothing—and their production and distribution entail practical considerations apart from normative claims. While material considerations may not be germane in a debate about how to split a cake which sits between us at arm's length, there are certainly many norms which we could agree to that could easily fail to be

practically or materially realizable (e.g. we should each get 55% of the cake). The point is simply that there are constraints on the norms that can be practically realized which are not determined by the discursive participants. That the nature of the material world has implications for politics is hardly a revelation, but discourse theories have no simple way to address this fact.

The general way of handling the material world in such discourse theories is to say that experts, scientific and technological elites, are responsible for debating and arriving at rational norms about the physical and material properties of the world. This, however, leaves non-experts ill-equipped to challenge such norms, and thus limits their ability to participate in a discourse based on those norms. But if we take such an approach, how are we ever going to be able to provide a useful critical theory of technology? Technological elites will always have a disproportionate amount of power in debates about technology, and to exclude technological elites or disregard the norms they have established is to ignore what is known about existing and potential technologies. What Winner hopes is that merely bringing technocrats and the masses together in participatory design will promote discourse, but this fails to address how the inherent power differential between the technocrats and the masses can be overcome.

While I believe that this techno-populism is a worthy objective, if only because it brings together people of diverse perspective and purpose to engage in the application of technologies, the promotion of such highly abstract ideals often contributes little more than rhetoric and motive to actual design projects. Moreover, it stresses democratic participation in technological choice as the principal lesson to be learned from participatory design—but it is not clear that it is practical or desirable to have universal participation in design choice. As was mentioned in the previous section, like many modern nations, participatory design only offers representational democracy, and thus introduces the possibility of misrepresentation. It would be unnecessary, or even impossible, for everyone who will ever ride a bus to participate in the design of that bus, yet it seems highly desirable to have meetings in which the bus-riding public voices their concerns over routes and schedules. The point is not that everyone gets a voice, but that everyone who has engaged the technology and is in a position to assess its usefulness in their daily practices has the ear of those who have the power to alter its potential usefulness. Participatory design researchers themselves stress the virtue of participation, but much of the value of their contribution lies in the consequences of realizing participation—the confrontation of the material and practical implications of their technological artifacts.

A critical theory of technology must include technology as an agent to be dealt with in political discourse. Unlike human political agents, technology does not challenge norms as illegitimate, but as unworkable. There is a danger of confusing this notion with a common assumption of techno-rationalism—that the technology itself imposes a logic on discourse. This is not what is meant by considering technology as an agent. Technologies are certainly open both to application and development, and are thus not inflexible to the political will. But neither are they infinitely flexible and perfectly plastic. Some configurations work, some cannot work, and many others may remain uncertain. Specific technologies may or may not be possible, and the only way to arrive at a legitimate conclusion is to pursue research, and thus commit resources, into exploring those possibilities. The material dialectic process can be most clearly seen in Engineering Codevelopment's “tuning” of an artifact to its environment. Such technological research is then a double dialectic: between political agents who determine the direction and goals of research, and between researchers and the material and practical resistances of specific technologies. What participatory design does is to add a third and shorter feedback loop to these interpenetrating dialectics. At the beginning of our history, these dialectic loops only touched at the points of project proposals and final system evaluation. Over time, the points of contact increased and the loops wound together. At the end of this history there are three dialectics: (1) users engage directly in a dialectic with the material and practical implications of a technological design, (2) enabling them to reformulate their desires and objectives within the dialectic between designers and users of the technology, (3) which in turn motivates designers in their dialectic engagement with the technology. Feenberg and Winner are concerned almost exclusively with (2). A critical theory of technology ought to recognize that ensuring an environment in which a public discourse occurs demands informed users and responsive technicians, and thus requires

dialectics (1) and (3). It is only through a direct engagement with the technology that one can decide whether a design can be realized, or whether the technology has satisfied the needs which motivate its construction.

Participatory design is an intriguing critical project which has successfully crossed multiple disciplinary, organizational, political and cultural boundaries. It traces the leading edges of technological application, business management, and social science and for this reason alone should be looked at carefully for its broader historical ramifications. But above all, participatory design has demonstrated that the critical engagement of technology requires not only a great deal of thoughtful reflection in the confrontation of political and ethical demands, but also that critical practice must fill the difficult role of articulating the technological alternatives which are actually capable of satisfying those demands.

References

1. Agre, P. (1995). From high tech to human tech: empowerment, measurement, and social studies of computing. Computer Supported Cooperative Work, 3, 167–195.
2. Anderson, W., & Crocca, W. (1993). Engineering practice and codevelopment of product prototypes. Communications of the ACM, 36 (4), 49–56.
3. August, J. (1991). Joint application design: the group session approach to system design. Englewood Cliffs, NJ: Yourdon Press.
4. Beyer, H., & Holtzblatt, K. (1995). Apprenticing with the customer. Communications of the ACM, 38 (5), 45–52.
5. Bion, W. (1946). The leaderless group project. Bulletin of the Menninger Clinic, 10, 77–81.
6. Braverman, H. (1974). Labor and monopoly capital: the degradation of work in the twentieth century. New York, NY: Monthly Review Press.
7. Brun-Cottan, F., & Wall, P. (1995). Using video to re-present the user. Communications of the ACM, 38 (5), 61–71.
8. Carmel, E., Whitaker, R., & George, J. (1993). PD and joint application design: a transatlantic comparison. Communications of the ACM, 36 (4), 40–48.
9. Churns, A. (1976). The principles of socio-technical design. Human Relations, 29, 783–904.
10. Clement, A., & Van den Besselaar, P. (1993). A retrospective look at PD projects. Communications of the ACM, 36 (4), 29–37.
11. Clifford, J. (1986). Introduction: partial truths. In J. Clifford, & G. Marcus, Writing culture: the poetics and politics of ethnography. Berkeley, CA: The University of California Press.
12. Crawford, A. (1994). Advancing business concepts in a JAD workshop setting: business reengineering and process redesign. Englewood Cliffs, NJ: Yourdon Press.
13. Ehn, P., & Kyng, M. (1987). The collective resource approach to systems design. In G. Bjerknes, P. Ehn, & M. Kyng, Computers and democracy: a Scandinavian challenge (pp. 17–57). Brookfield, VT: Avebury.
14. Emery, F., & Thorsrud, E. (1976). Democracy at work: the report of the Norwegian industrial democracy program. Leiden, Norway: Martinus Nijhoff.
15. Feenberg, A. (1991). Critical theory of technology. Oxford, UK: Oxford University Press.
16. Greenbaum, J. (1979). In the name of efficiency: management theory and shopfloor practice in dataprocessing work. Philadelphia, PA: Temple University Press.
17. Greenbaum, J., & Kyng, M. (1991). Design at work: cooperative design of computer systems. Hillsdale, NJ: Erlbaum.
18. Habermas, J. (1984). Theory of communicative action, vols 1 and 2. Boston, MA: Beacon Press (T. McCarthy, Trans.).
19. Habermas, J. (1990). The hermaneutic claim to universality. In G. Ormiston, & A. D. Shrift, The hermeneutic tradition from Aft to Ricoeur. Albany, NY: SUNY Press.
20. Hacking, I. (1983). Representing and intervening: introductory topics in the philosophy of natural science. Cambridge, UK: Cambridge University Press.
21. Hill, P. (1971). Towards a new philosophy of management. New York, NY: Gower.
22. Hirschheim, R., Klein, H. K., & Lyytinen, K. (1995). Information systems development and data modeling: conceptual and philosophical foundations. Cambridge, UK: Cambridge University Press.
23. In K. Holtzblatt, & H. R. Beyer (1995). Requirements gather: the human factor [special issue]. Communications of the ACM, 38 (5), 72–80.
24. Hutchings, A., & Knox, S. (1995). Creating products: customer demand. Communications of the ACM, 38 (5), 72–80.
25. Latour, B., & Woolgar, S. (1979). Laboratory life: the construction of scientific facts. Princeton, NJ: Princeton University Press.
26. Miller, P., & O'Leary, T. (1994). Accounting, "economic citizenship" and the spatial reordering of manufacture. Accounting, Organizations and Society, 19 (1), 15–43.
27. Miller, P., & Rose, N. (1995). Production, identity and democracy. Theory and Society, 24, 427–467.
28. Mumford, E. (1987). The collective resource approach to systems design. In G. Bjerknes, P. Ehn, & M. Kyng, Sociotechnical systems design: evolving theory and practice (pp. 59–76). Brookfield, VT: Avebury.
29. Mumford, E., & Henshall, D. (1979). A participative approach to computer systems design. London, UK: Associated Business Press.
30. Naisbitt, J., & Aburdene, P. (1985). Re-inventing the corporation: transforming your job and your company for the new information society. New York, NY: Warner Books.
31. Noble, D. (1977). America by design: science, technology, and the rise of corporate capitalism. Oxford, UK: Oxford University Press.
32. Noble, D. (1979). Social choice in machine design. In A. Zimbalist, Case studies on the labor process. New York, NY: Monthly Review Press.
33. Nurminen, M., & Weir, G. (1991). Computer jobs and human interfaces. Proceedings of the IFIP TC9/WG9.1 Working Conference. Amsterdam: North-Holland.
34. Pickering, A. (1995). The mangle of practice: time, agency and science. Chicago, IL: University of Chicago Press.
35. Sachs, P. (1995). Transforming work: collaboration, learning and design. Communications of the ACM, 38 (9), 36–44.
36. Schmidt, W., & Finnigan, J. (1992). The race without a finish line: America's quest for total quality. San Francisco, CA: Jossey-Bass Publishers.
37. Star, S. L. (1989). The structure of ill-structured solutions: heterogeneous problem-solving, boundary objects and distributed artificial intelligence. In M. Hahns, & L. Gasser, Distributed artificial intelligence, vol. 2 (pp. 37–54). Menlo Park, CA: Morgan Kauffman.
38. Suchman, L. (1995). Making work visible. Communications of the ACM, 38 (9), 56–64.
39. Trist, E. (1993). Introduction to volume II. In E. Trist, & H. Murray, The social engagement of social science: a Tavistock anthology. The socio-technical perspective, vol. II. Philadelphia, PA: University of Pennsylvania Press.

40. Willmott, H. (1995). Process reengineering, information technology and the transformation of accountability: the remaindering of the human resource? In E. Wray-Bliss, & W. Orlikowski et al., Information technology and changes in organizational work. Proceedings of the IFIP WG8.2 Working Conference (pp. 62–88). London, UK: Chapman and Hall.
41. Winner, L. (1995). Citizen virtues in a technological order. In A. Feenberg, & A. Hannay, Technology and the politics of knowledge (pp. 65–84). Bloomington, IN: Indiana Press.
42. Winner, L. (1977). Autonomous technology: technics-out-of-control as a political theme. Cambridge, MA: MIT Press.
43. Wood, J., & Silver, D. (1989). Joint Application DesignÒ: how to design quality systems in 40% less time. New York, NY: John Wiley & Sons.
44. Zuboff, S. (1988). In the age of the smart machine: the future of work and power. New York, NY: Basic Books.

Forecasting the Price of Gold: An Error Correction Approach

Kausik Gangopadhyay [1], Abhishek Jangir [2], Rudra Sensarma [1]

[1] Indian Institute of Management Kozhikode, IIMK Campus P.O., Kozhikode 673570, India

[2] NAV Capital LLP, A-402, Phase 4, Lake Homes, Chandivali, Powai, Mumbai 400076, India

* Corresponding Author. E-mail: asaro@uiuc.edu

Abstract. Gold prices in the Indian market may be influenced by a multitude of factors such as the value of gold in investment decisions, as an inflation hedge, and in consumption motives. We develop a model to explain and forecast gold prices in India, using a vector error correction model. We identify investment decision and inflation hedge as prime movers of the data. We also present out-of-sample forecasts of our model and the related properties.

Keywords: Gold price; Cointegration; Vector error correction model; Inflation hedge.

1. Introduction

India is one of the major gold consuming countries in the world and high demand from India is acknowledged to be a major factor in determining international gold prices. High import demand is also cited as the primary reason for the country's persistent current account deficit. To the best of our knowledge, there is little previous research about what determines gold prices in India. Understanding the determinants of gold price will help in developing a predictive model for forecasting future prices. This can be useful for the purpose of portfolio decision-making of investors and also as a critical input for policy making.

We submit that this paper is the first of its kind to develop a model for explaining and forecasting gold prices in India. We estimate the nature of the relationship of gold price in India with key determinants such as the stock market index, oil prices, exchange rate, interest rate, and consumer price index (CPI). We find that gold is useful as a portfolio hedge as well as a hedge against inflation. Our model is able to predict future gold prices with reasonable levels of accuracy.

2. Research background

Among all precious metals, gold might be the most popular choice for investment. It has stood the test of time, and performed well during crisis situations such as market decline, currency failure, high inflation, war, and so on. It is regarded as a good hedge both against inflation as well as fall in value of other assets. The usefulness of gold as an inflation hedge would imply that when general prices are high, gold prices will also be high so that the asset can be sold in order to finance general spending activity. However the role of gold as a hedge against other assets (such as stocks, bonds, foreign currency) would mean that when the prices of other assets fall, the price of gold rises such that the resulting portfolio is diversified.

Many studies have looked into the pattern of gold prices (see e.g. Capie et al, 2005, Worthington, Pahlavani, 2007 and Baur, Lucey, 2010) to identify the factors that influence gold prices. Some of the factors that influence gold prices include inflation, exchange rate, bond prices, market performance, seasonality, income, oil prices, and business cycles. However, to the best of our knowledge, there is no work that has been done to examine gold prices in India.

We carry out an analysis to study the factors influencing gold prices in India by collecting monthly data on gold prices and other factors over a long time period. While the hedge factors are expected to work in India as in other countries, there is an additional role of gold that may not be relevant elsewhere and has been hitherto ignored in literature. Indians buy gold not just for investment but also for personal reasons, to be used as a luxury good (to wear as jewellery, to gift in weddings, for religious reasons and so on). If this reason to buy gold is significant, then higher affordability should

lead to increased demand and therefore higher price for gold. We capture the wealth effect through the stock market index.

The time series variables that we study are, largely, non-stationary variables. Therefore, we need to analyse them in a cointegrating framework. We use a vector error correction approach to model and forecast the price of gold. Our benchmark estimates are for the period April 1990–August 2013.

We find that gold price has a cointegrating relationship with the stock market index, exchange rate, CPI, US bond rates, and oil price. The stock market index has a negative relationship with gold price, contradicting the argument for gold being a luxury good but supporting the role of gold as a portfolio hedge. This is consistent with Baur and McDermott (2010). The exchange rate has a negative relationship with gold price implying that a stronger rupee is associated with costlier gold. Our finding demonstrates that gold is a good hedge against the dollar from the point of view of domestic investors, which is also the case for developed countries (Reboredo, 2013a).

Oil price has a negative relationship with gold price implying that gold is a good hedge against oil as an investment, in contrast to existing evidence from developed countries (see for example, Reboredo, 2013b). The CPI has a positive relationship with gold indicating that gold is a good inflation hedge, a result that has been previously obtained for developed countries (Ghosh et al, 2004 and Worthington, Pahlavani, 2007). Finally, US bond rates are negatively related to gold price, indicating that when returns from international investments fall, investors may switch to gold.

We tested for robustness of the results of our exercise. We have taken some commonly used transformations of the variables, for example, the logarithmic one. We have added difference polynomials of independent variables. Our findings are quite robust to these alternative specifications. The relationship established by us provides interesting insights into the role of gold in portfolio diversification and as a hedge against inflation in the Indian context. The predictive capacity of our error correction model beats alternative specifications such as the random walk, using different sub-periods, and forecasting horizons.

3. Data and methodology

Data source

The gold price data are obtained from the Reserve Bank of India's website. It is taken in real terms by deflating it, using the CPI. The CPI data we use are for urban non-manual employees and later for industrial workers maintained by the Labour Bureau, Government of India.1 We have taken the equity market index Sensex as a proxy for the stock market. Whenever Sensex suffers a decline, the loss stricken investors may move towards gold, which increases the demand for gold, which in turn increases the price of gold. On the other hand, if Sensex represents the wealth of the people, then a higher value of the Sensex may indicate that the purchasing power of people increases, so they may be able to afford more gold, whose price increases. Sensex data are obtained from the website of the Bombay Stock Exchange.2

When the exchange rate increases, it makes gold imports more expensive, leading to an increase in the domestic price of gold. The US Dollar–Indian Rupee (USD–INR) exchange rate is collected from Indexmundi website3 and Bloomberg. However if gold were a good hedge against the exchange rate then we would expect gold prices to be negatively related with the exchange rate. This would mean that a fall in the dollar value would induce investors to move towards gold thereby leading to higher gold prices.

When oil prices increase, then the cost of production increases which reduces the profits of investors who then switch to gold for safety. Therefore, oil prices are expected to have a positive effect on gold prices. But people do not buy oil simply as a factor of production; many trade it as a commodity for capital gains. So an increase in oil prices would be beneficial for such investors and they would not invest in gold. Again, this means a negative relationship between gold prices and oil prices. (Oil prices are obtained from the Indexmundi website).

The Indian bond market is still in its nascent stage. A select group of authorised domestic financial institutions are the only players in the secondary market and liquidity is limited across maturities. Therefore, the effect of the bond rates was not analysed; another reason is the lack of data on the bond

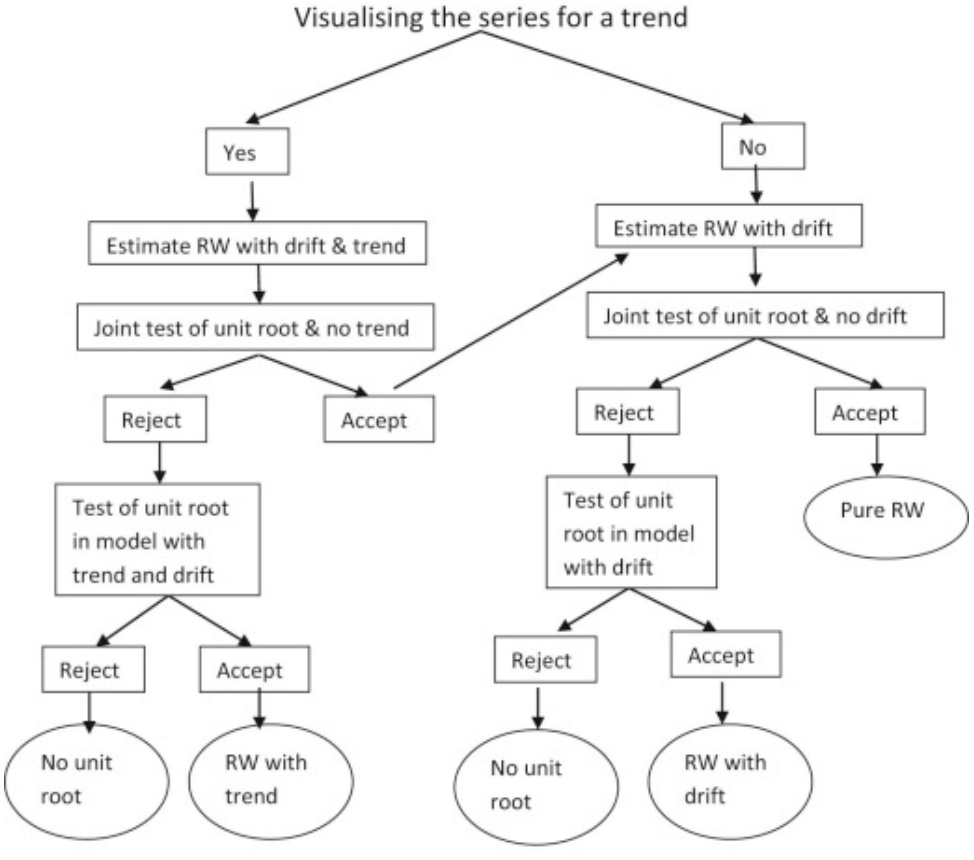
market, which was available only from 2004 after the setting up of the Clearing Corporation of India Ltd (CCIL). However we used interest rates on US bonds to control for international investment prospects and the data are obtained from the US treasury website. We use monthly data on the above variables between April 1990 and August 2013.

Data transformations

We consider real price of gold (GLP) which is free from the influence of general price movements. We have normalised gold price by dividing the nominal value by the consumer price index. For example, if the gold price in a particular month is INR 4508.91 and CPI is 2093, then GLP is calculated as INR 4508.91/2093. The Bombay Stock Exchange Sensitive Index or Sensex (SNX) is taken in a logarithmic scale to represent stock prices. Other determinants of gold price that we consider are the USD–INR exchange rate in logarithmic scale (denoted by EXR). We consider CPI in a logarithmic scale as well as the oil price (OIL). Finally we consider US bond rates (INT) to capture returns from international investments.

Pre-testing time-series properties of the data

We start by performing unit root tests for all our time series. There are several specifications of a unit root process: random walk, random walk with drift, random walk with linear trend and drift, and so on. The cookbook procedure for carrying out unit root test is schematically shown in Fig. 1. This procedure allows one not only to test for potential non-stationarity of the process but also to categorise the extent of the random walk process. An important practical issue for the implementation of the augmented Dickey–Fuller (ADF) test is the specification of the lag length p. If p is too small then the remaining serial correlation in the errors will bias the test. If p is too large then the power of the test will suffer.



RW: random walk

Fig.1. Schematic diagram for a cookbook procedure on testing for unit root.

For an optimal selection of lag length, we follow the procedure suggested by Ng and Perron (1995). We set an upper bound for p and estimate the ADF test regression. If the absolute value of the t-statistic for testing the significance of the last lagged difference is not significant then we decrease p

by one and repeat the same process; else we stop at that p. Schwert (1989) suggested the rule of thumb for determining the maximum lag length which is the highest integer contained in $12 \times (T/100)0.25$.

Once we have found some variables as non-stationary, we can find the long run relationship between detection of cointegrating relationship. Johansen's method of vector error correction model (VECM) is appropriate in this connection.

4. Econometric methodology

Cointegration

A set of variables are cointegrated when there exists a stable long run relationship between them. While the original test of cointegration was provided by Engle and Granger (1987), due to the well-known deficiencies of this simple approach (Enders, 2004), we follow the approach subsequently provided by Johansen and Juselius (Johansen, 1988, Johansen, 1991 and Johansen, Juselius, 1990). Formally put, let y1, y2, … , yk be a set of variables which we are interested in. Suppose each variable is integrated of order one, viz. I(1), there is a need for differencing in order to attain stationarity. If there exists linear combination(s) of the variables which is (are) I(0), then the variables are said to be cointegrated, i.e. they have a stable long run relationship. Then the cointegrating vector can be estimated which quantifies the relationship between the concerned variables.

Vector error correction model

The vector error correction model (VECM) involves expressing an nx1 vector of stationary time series (say yt) in terms of a constant, lagged values of itself and an error correction term. The standard VECM (p) model can be represented as,

where ECT refers to the Error Correction Term that is a product of an adjustment factor (α) and the cointegrating vector (β). The cointegrating vector shows the long term equilibrium relationship between the concerned variables while the adjustment factors show the speed of adjustment towards equilibrium in case there is any deviation.

5. Findings

Summary of data

Fig. 2 plots all the relevant variables over time. The results of unit root tests are summarised in Table 1. The unit root tests clearly indicate that the relevant variables are integrated of order 1, which is indicative of the most elementary degree of non-stationarity.

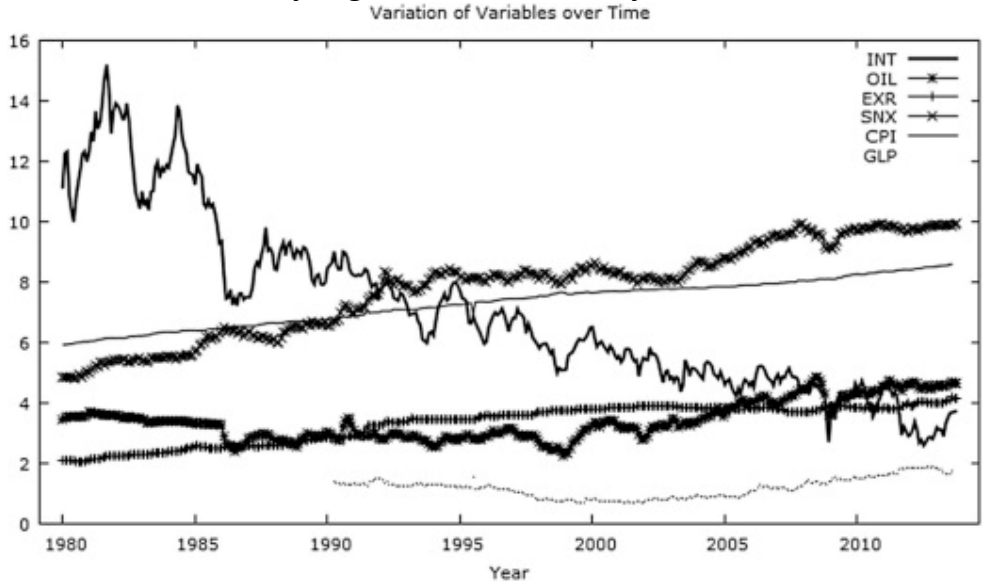


Fig.2. Variables considered in our analysis are plotted against time. The various series from top to bottom in the legend are INT: US bond rates, OIL: oil price, EXR: USD–INR exchange rate in logarithmic scale, SNX: Sensex, CPI: Consumer price index, GLP: price of gold.

Results from modelling gold price

The results from the test of cointegration are reported in Table 2. The trace test suggests that ranks of 0 and 1 are rejected at 5% level of significance. However rank of 2 cannot be rejected at the 5% level. In other words we can conclude that the variables have one cointegrating relationship among them.

Once the presence of cointegration is established we move to estimation of the cointegrating vectors and the VECM. Table 3 shows the cointegrating vector along with the standard errors of the estimates in parentheses. The coefficients suggest a relationship of the following nature:

The above equation indicates that gold prices and the stock market move in the opposite direction in the long run. Unlike what is expected of a luxury good, the wealth effect does not seem to dominate in the sense that higher wealth (captured by a rise in the Sensex) does not get reflected in increased demand and price of gold. However the role of gold as a hedge clearly dominates. As gold price moves in a different direction to that of the stock market it may be inferred that the role of gold as a portfolio hedge dominates its use as a luxury good in India.

This result may also imply that gold is a safe haven asset—an asset which investors can move into in times of high volatility. The difference between a hedge and a safe haven lies in the holding of the alternative asset under extreme market conditions. Under extreme market fluctuations investors tend to hold a safe haven asset whereas a hedge is a substitute for another asset when the latter is not performing as well.

Exchange rate is negatively related to gold prices in spite of the fact that the bulk of the gold consumed in India is imported. Therefore a weaker exchange rate should translate into higher cost of imported gold, which would also make domestic price higher. However our result seems to indicate that gold is a hedge against the dollar and could also be a safe haven. US bond rates have a positive relationship with gold which suggests that when returns from investing outside the country are high, demand for gold in India may fall, and therefore its price. Oil price has a positive relationship implying that gold may act as a good hedge against prices of commodities such as oil that are held by investors in their portfolio.

Forecasting the price of gold

To test our model's out-of-sample properties, we re-estimate our model restricting our data to August 2012. We use our model to forecast one year of data (Fig. 3) and find that the average error is about 3.47% (see Table 4). The root mean squared error of 0.11 is much less than the equivalent measure for a random walk model which is 0.75.

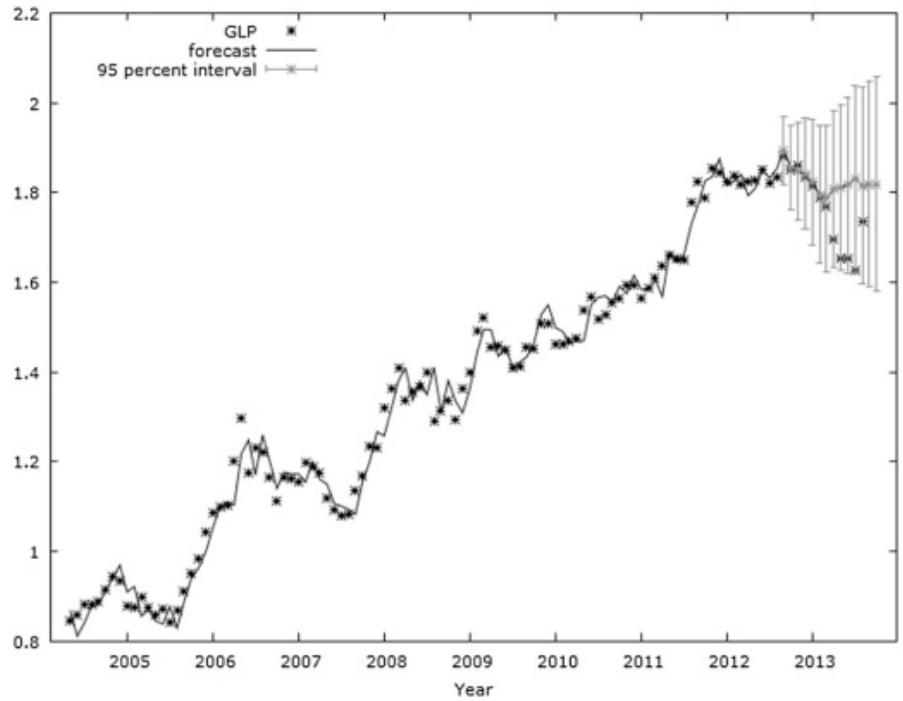
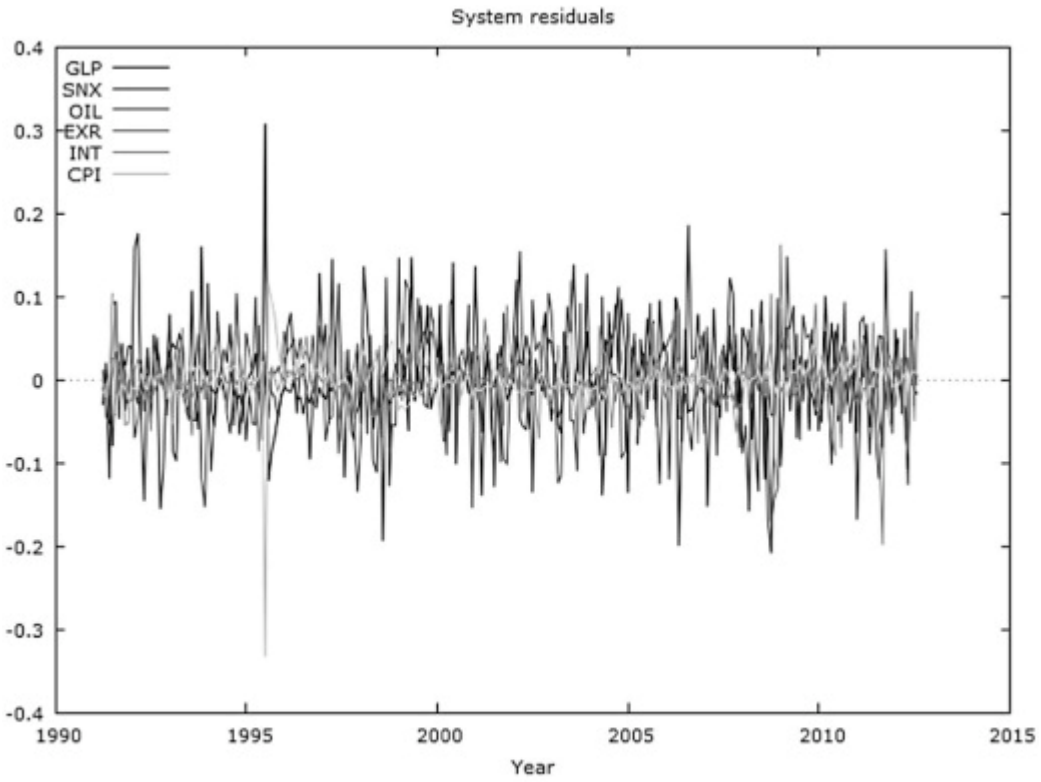


Fig.3. Forecast values: The magnitude of error. GLP: price of gold.

Robustness exercises

We carry out a series of exercises to assess the robustness of our results. For our model to be reliable, the estimated error terms should be normally distributed without any significant autocorrelation. This can be done using the Ljung–Box test that has good small sample properties as well. The null hypothesis of this test assumes the error terms to be normally distributed and independent. Hence an acceptance of the null would imply that our estimates are robust. We carry out the test for the estimated error term from all the equations of the VECM. For all the equations the p-values of the test statistics are actually very high (>0.90) which we interpret as non-existence of serial autocorrelation. The combined residual plots (Fig. 4) bear testimony to our analysis. Furthermore we tried out the above exercise with alternative lag orders from 9 to 15 (benchmark model presented is for lag order 12) but the results thus obtained are qualitatively not different from the benchmark model.



INT: US bond rates, OIL: oil price, EXR: USD-INR exchange rate in logarithmic scale, SNX: Sensex, CPI: Consumer price index, GLP: price of gold.

Fig. 4. Forecast values: The magnitude of error.

6. Conclusion

We have modelled gold prices in India and shown it to have a long term relationship with the stock market index, exchange rate, US bond rates, oil prices and the consumer price index. We found evidence that the role of gold as a portfolio hedge dominates its use as a luxury good in India. Gold prices are negatively related with oil prices, further indicating the role of gold as a hedge. Gold prices go up when the rupee is weaker implying that gold is a good hedge against the dollar. When returns from investing outside the country are high, gold price in India is low. Finally, gold acts as a good inflation hedge as it moves in the same direction as CPI. We found evidence that the above variables are able to forecast gold over a 12-month horizon better than a random walk model.

One implication of our results is that since gold seems to be a useful portfolio hedge as well as inflation hedge, government policies to curb the import of gold may be futile. Yet the large amounts of gold imports are a cause for concern as they have kept India's current account deficit high leading to pressure on the rupee. Our research suggests that policies that directly address the causes of

inflation and provide alternative investment opportunities for retail investors may better serve the objective of bringing down gold imports.

Future work in this area can proceed in several directions. In terms of methodology, alternative approaches such as copula, artificial neural networks, Fourier transformation and wavelet analysis can be employed to assess the scope for improvement in the forecasting power. Other research approaches such as behavioural finance models can be tested using micro data on investors' personal choices to study their influence on gold prices. Studies can compare the gold holding decisions of households and corporate houses to evaluate the consumption vis-à-vis investment motives behind gold purchase.

References

1. Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. The Financial Review, 45(2), 217–229.
2. Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. Journal of Banking & Finance, 34(8), 1886–1898.
3. Capie, F., Mills, T. C., & Wood, G. (2005). Gold as a hedge against the dollar. Journal of International Financial Markets, Institutions and Money, 15(4), 343–352.
4. Enders, W. (2004). Applied econometric time series. Hoboken, NJ: John Wiley & Sons.
5. Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation, and testing. Econometrica: Journal of the Econometric Society, 55(2), 251–276.
6. Ghosh, D., Levin, E. J., Macmillan, P., & Wright, R. E. (2004). Gold as an inflation hedge? Studies in Economics and Finance, 22(1),1–25.
7. Johansen, S. (1988). Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control, 12(2–3), 231–254.
8. Johansen, S. (1991). A Bayesian perspective on inference from macroeconomic data: comment. Scandinavian Journal of Economics, 93(2), 249–251.
9. Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. Oxford Bulletin of Economics and Statistics, 52(2),169–210.
10. Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with datadependent methods for the selection of the truncation lag. Journal of the American Statistical Association, 90, 268–281.
11. Reboredo, J. C. (2013a). Is gold a safe haven or a hedge for the US dollar? Implications for risk management. Journal of Banking & Finance, 37(8), 2665–2676.
12. Reboredo, J. C. (2013b). Is gold a hedge or safe haven against oil price movements? Resources Policy, 38(2), 130–137.
13. Schwert, W. (1989). Test for unit roots: a Monte Carlo investigation. Journal of Business and Economic Statistics, 7, 147–159.
14. Worthington, A. C., & Pahlavani, M. (2007). Gold investment as an inflationary hedge: cointegration evidence with allowance for endogenous structural breaks. Applied Financial Economics Letters, 3(4), 259–262.

Adv. Res. Econ. Manage., 2016, 1(1): 54, C3.

Here is Your Paper's Title: Author Instructions

FULL First Author[1, a], FULL Second Author[2,b, *] and Others[3,c]

[1]Full address of first author, including country and zip code
[2]Full address of second author, including country and zip code
[3]List all distinct addresses in the same way
[a] email, [b] email, [c] email
* Corresponding Author: Prof. Name, PhD / M.D.. E-mail: abc@abc.com

Abstract. No more than 150 words. This document explains and demonstrates how to prepare your camera-ready manuscript for 1088.email Publications. The best is to read these instructions and follow the outline of this text. The text area for your manuscript must be 17 cm wide and 25 cm high (6.7 and 9.8 inches, resp.). Do not place any text outside this area. Use good quality, white paper of approximately 21 x 29 cm or 8 x 11 inches (please do not change the document setting from A4 to letter). Your manuscript will be reduced by approximately 20% by the publisher. Please keep this in mind when designing your figures and tables etc.

Keywords: keywords 1, keywords 2, keywords 3, keywords 4, keywords 5, keywords 6.

Introduction

All manuscripts must be in English. Please keep a second copy of your manuscript in your office (just in case anything gets lost in the mail). When receiving the manuscript, we assume that the corresponding authors grant us the copyright to use the manuscript for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper.

Use *italic* for emphasizing a word or phrase. Do not use boldface typing or capital letters except for section headings (cf. remarks on section headings, below). Use a laser printer, not a matrix dot printer.

Organization of the Text

Section Headings. The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph).

Page Numbers. Do *not* print page numbers: Please number each sheet toward the middle near the bottom (outside the typing area) with a soft pencil.

Tables. Tables (refer with: Table 1, Table 2, ...) should be presented as part of the text, but in such a way as to avoid confusion with the text. A descriptive title should be placed above each table. The caption should be self-contained and placed *below or beside* the table. Units in tables should be given in square brackets [meV]. If square brackets are not available, use curly {meV} or standard brackets (meV).

Special Signs. for example , $\alpha\ \gamma\ \mu\ \Omega\ ()\ \geq\ \pm\ \bullet\ \Gamma\ \{11\ \bar{2}\ 0\}$ should always be written in with the fonts Times New Roman or Arial

Figures. Figures (refer with: Fig. 1, Fig. 2, ...) also should be presented as part of the text, leaving enough space so that the caption will not be confused with the text. The caption should be self-contained and placed *below or beside* the figure. Generally, only original drawings or photographic reproductions are acceptable. Only very good photocopies are acceptable. Utmost care must be taken to *insert the figures in correct alignment with the text*. Half-tone pictures should be in the form of glossy prints. If possible, please include your figures as graphic images in the electronic version. For best quality the pictures should have a resolution of 300 dpi(dots per inch).

Color figures are welcome for the online version of the journal. Generally, these figures will be reduced to black and white for the print version. The author should indicate on the checklist if he wishes to have them printed in full color and make the necessary payments in advance.

Equations. Equations (refer with: Eq. 1, Eq. 2, ...) should be indented 5 mm (0.2"). There should be one line of space above the equation and one line of space below it before the text continues. The equations have to be numbered sequentially, and the number put in parentheses at the right-hand edge of the text. Equations should be punctuated as if they were an ordinary part of the text. Punctuation appears after the equation but before the equation number, e.g.

$$c^2 = a^2 + b^2. \quad (1)$$

Literature References

References are cited in the text just by square brackets [1]. (If square brackets are not available, slashes may be used instead, e.g. /2/.) Two or more references at a time may be put in one set of brackets [3,4]. The references are to be numbered in the order in which they are cited in the text and are to be listed at the end of the contribution under a heading *References*, see our example below.

Summary

On your CD, please indicate the format and word processor used. Please also provide your phone number, fax number and e- mail address for rapid communication with the publisher. Please *always* send your CD along with a hard copy that must match the CD's content *exactly*. If you follow the foregoing, your paper will conform to the requirements of the publisher and facilitate a problem-free publication process.

Acknowledgement

For example: This paper is supported by the National Foundation (123456, 34567 and 56789), the Research Program of 1088.Email (2016SRWLV007) and the Foundation of 1088.Email (1088EP2016-1234). The authors would like to acknowledge the constructive comments given by the anonymous reviewers.

References

[1] Dj.M. Maric, P.F. Meier and S.K. Estreicher: Mater. Sci. Forum Vol. 83-87 (1992), p. 119

[2] M.A. Green: *High Efficiency Silicon Solar Cells* (Trans Tech Publications, Switzerland 1987).

[3] Y. Mishing, in: *Diffusion Processes in Advanced Technological Materials*, edtied by D. Gupta Noyes Publications/William Andrew Publising, Norwich, NY (2004), in press.

[4] G. Henkelman, G.Johannesson and H. Jónsson, in: Theoretical Methods in Condencsed Phase Chemistry, edited by S.D. Schwartz, volume 5 of Progress in Theoretical Chemistry and Physics, chapter, 10, Kluwer Academic Publishers (2000).

[5] R.J. Ong, J.T. Dawley and P.G. Clem: submitted to Journal of Materials Research (2003)

[6] P.G. Clem, M. Rodriguez, J.A. Voigt and C.S. Ashley, U.S. Patent 6,231,666. (2001)

[7] Information on http://www.weld.labs.gov.cn

...

Advanced Research on Economics and Management
ISSN Pending
Volume 1 number 1
May 1, 2016
Published on *http://www.1088.email*