On the Use of Cross-Sectional Measures of Forecast Uncertainty

Ciaran Driver
(School of Oriental and African Studies, UK)

Lorenzo Trapani*
(Cass Business School, UK)

Giovanni Urga
(Cass Business School, UK, and University of Bergamo, Italy)

**Corresponding author*: Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ (U.K.). Tel. 00.44.(0)20.70405260 Fax. 00.44.(0)20.70408881, e-mail: L.Trapani@city.ac.uk

On the Use of Cross-Sectional Measures of Uncertainty

Abstract. This paper investigates the role of cross sectional dependence among private forecasters, assessing its impact on measuring and using forecasting uncertainty. We study under which circumstances cross sectional measures of uncertainty (such as disagreement across forecasters) are valid proxies for private information, analysing the impact of distributional assumptions on private signals. In particular, we explore the role played by cross dependence among forecasters, arising e.g. from partially shared private information. We validate the theory through a Monte Carlo exercise, which reinforces our findings, and through an application to US nonfarm payroll data.
J.E.L. Classification Numbers: C21, C22.
Keywords: Forecast Disagreement, Cross Sectional Dependence, Uncertainty.

1 Introduction

The main question of this paper is: can disagreement among private forecasters (irrespective of its determinants) be used to improve predictive ability? We base our analysis on the same setup as in Engle (1983), using a simple model in which an outcome variable, y_t, has DGP determined by its past value(s) and some other explanatory variables, which may be observable or not. In this context, y_t is predicted by a researcher by using only past information. This can be due to the other explanatory variables being unobservable, or to the model (s)he employs simply not including them. As well as by the researcher, y_t may be predicted by several individuals, who use the publicly available past information on y_t and some other explanatory variables. Such variables may be available only to them; alternatively, the variable may be observable but only used by some forecasters. Based on this framework, we propose that the researcher, in addition to using standard proxies such as the mean of the individual forecasts, should proxy the unavailable explanatory variables by using a measure of dispersion among the individual forecasters. The most obvious measure of dispersion is the cross sectional variance (henceforth defined as $CS_t^{(2)}$), which is traditionally used as a measure of disagreement (see e.g. Giordani and Soderlind, 2003). We show that using $CS_t^{(2)}$ as a regressor in a model for y_t can increase forecasting ability by reducing the Mean Squared Error (MSE) of forecasts. However, we also show that the usefulness of $CS_t^{(2)}$ is very sensitive to the distributional features of the explanatory variables in the DGP of y_t. Indeed, as a leading counterexample we show that augmenting an ARMA specification for y_t by including $CS_t^{(2)}$ yields no gain in predictive ability when the omitted explanatory variables follow a normal distribution. In order to generalise this, we also consider a generalised version of cross sectional disagreement, called $CS_t^{(k)}$, which, in essence, is based on representing cross sectional dispersion as the k-th sample moment of individual forecasts. Such generalised measures are not sensitive to the distribution of the omitted variables. Given that the $CS_t^{(k)}$s are non structural in nature, the approach that we recommend is a General-to-Specific (GETS) approach based on using the AutoMetrics option on OxMetrics 6.2 (Doornik, 2009; Castle et al., 2011), fitting an ADL model to y_t using the $CS_t^{(k)}$s, and their lags.

We show that the $CS_t^{(k)}$s manage to proxy the extra omitted variables by exploiting the presence of cross dependence among them. This has a twofold implication. On the one hand, cross dependence among the information sets available to individual forecasters is necessary in order for the $CS_t^{(k)}$s to improve forecasting ability. On the other hand, the reverse argument holds: whenever there are cross dependent forecasts, even in presence of unobservable, private information, it is possible to proxy such private information by using disagreement, and use it to better predict y_t. Individual forecasts that are correlated have been noted in various contexts. Examples include the accounting literature, where evidence of correlation among individual earnings forecasts has been found in several contributions (e.g. O'Brien, 1988; Lys and Sohn, 1990; see also the analysis in Fischer and Verrecchia, 1998; Barron et al. 1998); macroeconomics, using surveys of professional forecasters (Dovern et al., 2011; Genre et al., 2010); in predicting unemployment, using the Blue Chips Survey (Gregory et al., 2001); and we also refer to the comments on the presence of correlation in the Survey of Professional Forecasters in Elliott (2011), and the theoretical framework therein, where the impact of cross correlation in determining the optimal forecast is discussed. Thus, cross dependence among forecasters is an important feature in empirical studies. Several theoretical explanations have been proposed, from presence of partially shared private information (Patton and Timmermann, 2010) to herding (Scharfstein and Stein, 1990; Stein, 2003).

As well as the contribution above, we review the relationship between the traditionally employed measure of cross sectional dispersion $CS_t^{(2)}$, and GARCH-type measures. As mentioned above, this is a "classical" investigation (see Lahiri and Sheng, 2010); in our context, we assess the impact of cross dependence on this relationship, showing that it leads to an ambiguous sign in the difference between them.

The paper is organised as follows. We first (Section 2) introduce the $CS_t^{(k)}$s, showing how they can reduce the forecast error for y_t. This is validated through a Monte Carlo exercise and through the application to the prediction of the United States Non-farm Payroll index (NFP henceforth) - Section 3. Section 4 concludes. All proofs and derivations are in the supplementary online material.

2 Use of cross section uncertainty in forecasting

The starting point of our analysis is equation (9) in Engle (1983, p. 295):

$$y_t = \beta y_{t-1} + \alpha_t' \varepsilon_t + \eta_t, \tag{1}$$

with $|\beta| < 1$. Equation (1) states that y_t is generated by a process which depends on its past value(s), and on a set of n explanatory variables, $\varepsilon_t \equiv [\varepsilon_{1t}, ..., \varepsilon_{nt}]'$; this could be regarded as "the reduced form of a structural model" (Engle, 1983, p. 295), with η_t being the error term. As far as forecasting y_t is concerned, we start our analysis from the same viewpoint as Engle: y_t is predicted by a researcher who has only y_{t-1} at his/her disposal. Thus, the researcher predicts y_t as

$$y_t^r = \beta y_{t-1}.$$

Hence, the forecast error in this case is $y_t - y_t^r = \alpha_t' \varepsilon_t + \eta_t$. In this respect, (1), from the researcher's viewpoint, is a model with latent explanatory variables (the ε_{it}s). The researcher's model is

$$y_t = \beta y_{t-1} + v_t, \tag{2}$$

where $v_t = \eta_t + \alpha_t' \varepsilon_t$. Thus, from the researcher's viewpoint, using (2) instead of (1) is an omitted variables problem.

Alongside the researcher, Engle's framework postulates the existence of n forecasters, each of whom has inside information on his/her own ε_{it}; in this respect, ε_{it} is customarily interpreted as private information, but more generally it represents the additional regressors that the i-th forecasters uses in order to predict y_t. This entails that the i-th forecaster predicts y_t as

$$y_t^i = \beta y_{t-1} + \alpha_{it}\varepsilon_{it}. \tag{3}$$

Equation (3) is based on the assumption that β and α_{it} are observable, and therefore it may be viewed as an infeasible prediction. We use this as our baseline case. In the comments to Proposition 1 below, we analyse the impact of having to estimate both β and α_{it} on the prediction y_t^i.

We consider the following assumptions.

Assumption 1: η_t and ε_t are mutually independent, zero mean, covariance stationary processes with $E(y_{t-1}\eta_t) = 0$, $E(y_{t-1}\varepsilon_t) = 0$, $Var(\eta_t) = \sigma_\eta^2 < \infty$ and $E(\varepsilon_{it}\varepsilon_{jt}) = \omega_{ij}$ with $\omega_{ij} = 1$ for all $i = j$.

Assumption 2: the α_{it}s are non-stochastic quantities that satisfy *(i)* $\alpha_t'\varepsilon_t = O_p(1)$, *(ii)* $0 < \alpha_t' E(\varepsilon_t\varepsilon_t')\alpha_t < \infty$ as $n \to \infty$ for all t; *(iii)* $0 < \sum_{i=1}^n \alpha_{it}^2\varepsilon_{it}^2 < \infty$ as $n \to \infty$ for all t; *(iv)* $E\left[(\alpha_t'\varepsilon_t)^4\right] < \infty$ and $E\left[\left(\sum_{i=1}^n \alpha_{it}^2\varepsilon_{it}^2\right)^2\right] < \infty$ as $n \to \infty$; *(v)* $\|\alpha_t\| = O(1)$ for all n and t.

Assumption 1 considers the presence of contemporaneous correlation, and therefore of interactions among agents. As pointed out above, from the researcher's viewpoint, using (2) instead of (1) is an omitted variables problem; assuming $E[y_{t-1}\varepsilon_t] = 0$ entails that this does not cause inconsistency of the estimated β.

Assumption 2 allows the α_{it}s to be time dependent; this also entails that the number of forecasters, n, is allowed to vary over time, as it is typical in empirical applications. In addition to this, Assumption 2 poses some restrictions on the moments of $\alpha_t'\varepsilon_t$ as $n \to \infty$. The square summability condition prevents the variance of the error term in regression (1) from exploding as the number of individuals grows; a similar assumption is contained in Pesaran and Weale (2006).

The regressors ε_{it} are not observable to the researcher. Thus, (s)he could proxy them using some variables that are related to them. In order to construct such an "instrument", recall that each individual forecaster predicts y_t using y_{t-1} and ε_{it}. The i-th forecaster's prediction error is given by

$$\varepsilon_t^i \equiv y_t - y_t^i = \eta_t + \sum_{j\neq i} \alpha_{jt}\varepsilon_{jt}. \tag{4}$$

Define $CS_t^{(2)}$ as the dispersion of the individual predictions around their mean

$$\begin{aligned} CS_t^{(2)} &= \sum_{i=1}^n \left(y_t^i - \bar{y}_t\right)^2 \\ &= \sum_{i=1}^n \alpha_{it}^2\varepsilon_{it}^2 - \frac{1}{n}\left(\alpha_t'\varepsilon_t\right)^2, \end{aligned} \tag{5}$$

where the second equality follows from $y_t^i - \bar{y}_t = y_t^i - \frac{1}{n}\sum_{i=1}^n y_t^i = \alpha_{it}\varepsilon_{it} - \frac{1}{n}\alpha_t'\varepsilon_t$.

Equation (5) illustrates how $CS_t^{(2)}$ can be used by the researcher as a proxy for the ε_{it}s. The quantity $CS_t^{(2)}$ contains the squares of the ε_{it}s, and it is observable at time t, since it is constructed using predictions for y_t which are available prior to t. From a technical point of view, our definition of $CS_t^{(2)}$ is different from the one usually employed in the literature, where cross sectional dispersion is defined as $\frac{1}{n}\sum_{i=1}^{n}\left(y_t^i - \bar{y}_t\right)^2$. In our case, there is no need to divide by n, since $\sum_{i=1}^{n}\left(y_t^i - \bar{y}_t\right)^2$ is already normalised by assuming that the α_{it}s are summable. In order to make the two measures comparable, further assumptions are needed on the α_{it}s, e.g. that they sum to one.

Since $CS_t^{(2)}$ contains a transformation of the ε_{it}s, in order to reduce the error term in (2), i.e. $v_t = \eta_t + \alpha_t'\varepsilon_t$, the econometrician could employ the augmented regression:

$$y_t = \beta y_{t-1} + \gamma CS_t^{(2)} + v_t^*, \tag{6}$$

where

$$v_t^* = v_t^*\left(\gamma\right) = \eta_t + \alpha_t'\varepsilon_t - \gamma CS_t^{(2)}.$$

Considering an MSE criterion, using $CS_t^{(2)}$ improves the prediction of y_t in (2) as long as $E\left(v_t^{*2}\right) < E\left(v_t^2\right)$. Particularly, for the case of an estimator that minimizes $E\left(v_t^{*2}\right)$, the model improves after adding $CS_t^{(2)}$ if $\gamma \neq 0$ (otherwise there is only an unnecessary reduction in the degree of freedom) and if, for the chosen value of γ (say γ^*), $E\left(v_t^{*2}\right)$ is indeed smaller than $E\left(v_t^2\right)$.

It holds that:

Proposition 1 *Let Assumptions 1-2 hold with* $E\left\|\varepsilon_t\right\|^4 < \infty$ *and consider*

$$\min_{\gamma} E\left[v_t^*\left(\gamma\right)\right]^2. \tag{7}$$

This has solution

$$\begin{aligned}\gamma^* &= \frac{n^2 E\left[\alpha'\varepsilon_t \sum_{i=1}^{n}\alpha_{it}^2\varepsilon_{it}^2\right] - nE\left[\left(\alpha_t'\varepsilon_t\right)^3\right]}{E\left[\left(\alpha_t'\varepsilon_t\right)^2 - n\left(\sum_{i=1}^{n}\alpha_{it}^2\varepsilon_{it}^2\right)\right]^2} \\ &= \frac{E\left[CS_t^{(2)}\left(\alpha_t'\varepsilon_t\right)\right]}{E\left[\left(CS_t^{(2)}\right)^2\right]}.\end{aligned} \tag{8}$$

Also, it holds that $E\left[v_t^*\left(\gamma^*\right)\right]^2 \leq E\left[v_t^2\right]$, *with* $E\left[v_t^*\left(\gamma^*\right)\right]^2 = E\left[v_t^2\right]$ *if and only if* $\gamma^* = 0$. *The same result holds as* $n \to \infty$, *assuming that* $\sup_i E\left|\varepsilon_{it}\right|^4 < \infty$ *and* $\sup_i\left|\alpha_{it}\right| = O\left(n^{-1/4}\right)$.

Proposition 1 states that it is possible to attenuate the MSE by using disagreement among forecasters as an explanatory variable. This is accomplished by proxying the unobservable ε_{it}s using $CS_t^{(2)}$, which contains (a quadratic transformation of) the ε_{it}s; improvements are present when $\gamma^* \neq 0$.

It is interesting to explore what happens when the number of forecasters or alternative models, n, passes to infinity. As $n \to \infty$, Assumption 2 ensures that $\alpha_t' \varepsilon_t$ and $\sum_{i=1}^n \alpha_{it}^2 \varepsilon_{it}^2$ do not vanish. Thus, $CS_t^{(2)} = \sum_{i=1}^n \alpha_{it}^2 \varepsilon_{it}^2 + O_p\left(n^{-1}\right)$. From (8), this entails that $\gamma^* = O_p\left(1\right)$.

Building on the calculations in the supplementary material, it can be shown that

$$\begin{aligned} E\left[v_t^*\left(\gamma^*\right)\right]^2 - E\left[v_t^2\right] &= \gamma^{*2} E\left[\left(\sum_{i=1}^n \alpha_{it}^2 \varepsilon_{it}^2\right)^2\right] - 2\gamma^* E\left[\alpha_t' \varepsilon_t \sum_{i=1}^n \alpha_{it}^2 \varepsilon_{it}^2\right] \\ &\quad + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{n^2}\right). \end{aligned}$$

Thus, it follows from (8) that

$$E\left[v_t^*\left(\gamma^*\right)\right]^2 - E\left[v_t^2\right] = -\frac{\left[E\left(CS_t^{(2)} \alpha_t' \varepsilon_t\right)\right]^2}{E\left[\left(CS_t^{(2)}\right)^2\right]},$$

which is always negative as long as $E\left(CS_t^{(2)} \alpha_t' \varepsilon_t\right) \neq 0$. Therefore, as $n \to \infty$, there is still a gain in forecasting accuracy.

From a technical point of view, the main result in Proposition 1 (i.e., the possibility of proxying private information through $CS_t^{(2)}$) holds under more general conditions than Assumptions 1 and 2. For example, if we assumed that ε_t and η_t are not independent, equation (8) would be modified into $\gamma^* = \frac{E\left[CS_t^{(2)}\left(\alpha_t' \varepsilon_t - \eta_t\right)\right]}{E\left[\left(CS_t^{(2)}\right)\right]}$ - this follows from the same algebra as in the proof of the Proposition. Even in this case, γ^* is not, in general, equal to zero, and thus it can be employed as a proxy for the ε_{it}s.

It is worth pointing out that the result in Proposition 1 are based on the assumption that the i-th forecaster knows and uses the actual values of β and α_i (we suppress the dependence on i for simplicity). As it can be expected, Proposition 1 does not change if β and α_i are replaced with consistent estimators in (3). The i-th forecaster would estimate β and α_i from his/her model, viz.

$$y_{it} = \beta y_{t-1} + \alpha_i \varepsilon_{it} + v_t^i, \tag{9}$$

with $v_t^i = \eta_t + \sum_{j \neq i} \varepsilon_{it}$. In view of (1), this is an omitted variables problem, similarly to the one observed for the researcher when using (2). Indeed, estimating β consistently is possible for the i-th forecaster, due to the independence between y_{t-1} and the omitted ε_{it}s. However, consistent estimation of α_i from (9) is fraught with difficulties. Of course, if the ε_{it}s are assumed to be uncorrelated (as in Engle, 1983), then α_i can be estimated applying e.g. OLS to (9) and the estimate can be expected to be consistent. Conversely, if the ε_{it}s are correlated, consistency may not hold. The complete passages are in the supplementary online material; here, we briefly show that even using inconsistently estimated α_is does not make $CS_t^{(2)}$ useless. Consider (9), and, for simplicity, let $\beta = 0$ and assume that ε_{it} is normalised so that $\sum_{t=1}^{T} \varepsilon_{it}^2 = T$. The OLS estimation error of α_i is

$$\hat{\alpha}_i - \alpha_i = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} \eta_t + \sum_{j \neq i} \alpha_j \left(\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it} \varepsilon_{jt} \right) = I + II.$$

Considering I, this is $O_p\left(T^{-1/2}\right)$ under Assumption 1. Turning to II, the only way in which $\hat{\alpha}_i$ can be consistent is if $\sum_{j \neq i} \alpha_j \left(T^{-1} \sum_{t=1}^{T} \varepsilon_{it} \varepsilon_{jt} \right) = o_p(1)$. Using the notation in Assumption 1 and a LLN, this could be rewritten as $\sum_{j \neq i} \alpha_j \omega_{ij} = o_p(1)$. This holds e.g. if $\omega_{ij} = O\left(T^{-\nu}\right)$ for some $\nu > 0$ for all $j \neq i$, but this may be a rather artificial requirement. Thus, in general, $\hat{\alpha}_i$ is estimated inconsistently when using (9), which is not surprising due to the omitted variable problem mentioned above. From the researcher's point of view, this entails that $CS_t^{(2)}$ should be replaced by its empirical counterpart, say $\widehat{CS}_t^{(2)}$, defined as

$$\widehat{CS}_t^{(2)} = \sum_{i=1}^{n} \left(y_t^i - \bar{y}_t \right)^2 = \sum_{i=1}^{n} \hat{\alpha}_i^2 \varepsilon_{it}^2 - \frac{1}{n} \left(\hat{\alpha}' \varepsilon_t \right)^2,$$

so that (6) becomes

$$y_t = \beta y_{t-1} + \gamma \widehat{CS}_t^{(2)} + \hat{v}_t^*,$$

with $\hat{v}_t^* = \hat{v}_t^*(\gamma) = \eta_t + \alpha_t' \varepsilon_t - \gamma \widehat{CS}_t^{(2)}$. However, following the same passages as in the proof of Proposition 1, it can be shown that the solution to the minimisation problem $\min_\gamma E\left[\hat{v}_t^*(\gamma)\right]^2$ is $\left\{ E\left[\left(\widehat{CS}_t^{(2)} \right)^2 \right] \right\}^{-1} E\left[\widehat{CS}_t^{(2)} \left(\alpha' \varepsilon_t \right) \right]$, which is not, in general, equal to zero. Thus, from the researcher's point of view, using $\widehat{CS}_t^{(2)}$ does not, in general, cause problems, even if the α_is (or some of them) are not estimated consistently. The intuition behind this is that the estimation error $\hat{\alpha}_i - \alpha_i$ may not vanish as $T \to \infty$, but its asymptotic bias contains the ε_{jt}s (with $j \neq i$), which is indeed useful information. From an empirical point of view, of course the researcher does not know how the α_is have been estimated, and therefore the only way of assessing whether using $\widehat{CS}_t^{(2)}$ is useful is to estimate (6) and check whether γ is significantly different from zero.

Finally, note that equation (8) also illustrates some potential issues with using $CS_t^{(2)}$: in general, the usefulness of $CS_t^{(2)}$ (and of $\widehat{CS}_t^{(2)}$) depends in a non-trivial way on the (unobservable) distributional properties of the ε_{it}s. In order to illustrate this, we consider, as an example, the case of the ε_{it}s being Gaussian; assuming normality of private signals is a typical assumption in the literature (see e.g. the literature on herding: Chamley, 2004). In such case, it holds that $\gamma^* = 0$, and therefore $CS_t^{(2)}$ is not useful. This is due to the fact that in the numerator of (8) there are quantities like $E\left(\varepsilon_{it}^3\right)$ and $E\left(\varepsilon_{it}^2 \varepsilon_{jt}\right)$ with $i \neq j$, which are all equal to zero if ε_{it} is Gaussian. This is only an illustrative example, based on the infeasible i-th prediction y_t^i (see equation (3)), and in Section 3 we report a set of simulations under various distributional assumptions to analyse in which cases $CS_t^{(2)}$ can be employed.

In order to expand the framework and to make it robust to the distributional properties of the ε_{it}s, we introduce a generalised class of measures of cross sectional disagreement. Define the k-th sample moment of the individual forecasts:

$$\begin{aligned} CS_t^{*(k)} &= \sum_{i=1}^{n} \left(y_t^i - \bar{y}_t\right)^k \\ &= \sum_{i=1}^{n} \left(\frac{1}{n}\alpha_t'\varepsilon_t - \alpha_{it}\varepsilon_{it}\right)^k = \sum_{i=1}^{n}\sum_{j=0}^{k} \frac{1}{n^{k-j}}\binom{k}{j}\left(\alpha_t'\varepsilon_t\right)^{k-j}\alpha_{it}^j\varepsilon_{it}^j, \end{aligned} \tag{10}$$

for $k = 2, ..., p$, where the last equality comes from Pascal's triangle.

The definition of $CS_t^{*(k)}$ is based on the case of finite n. However, as $n \to \infty$, it is possible to envisage that $\sum_{i=1}^{n}\left(\frac{1}{n}\alpha_t'\varepsilon_t - \alpha_{it}\varepsilon_{it}\right)^k$ converges to zero. Indeed, using the C_r inequality, $\sum_{i=1}^{n}\left(\frac{1}{n}\alpha_t'\varepsilon_t - \alpha_{it}\varepsilon_{it}\right)^k \leq Cn^{1-k}\left(\alpha_t'\varepsilon_t\right)^k + C\sum_{i=1}^{n}\alpha_{it}^k\varepsilon_{it}^k$. Assumption 2(i) stipulates that $n^{1-k}\left(\alpha_t'\varepsilon_t\right)^k = O_p\left(n^{1-k}\right)$. The argument for $\sum_{i=1}^{n}\alpha_{it}^k\varepsilon_{it}^k$ is subtler, but again in light of Assumption 2(i), it is natural to think of the case of α_{it} being proportional to $n^{-1/2}$ (albeit non necessary; the assumption allows for greater flexibility). In such case, $\sum_{i=1}^{n}\alpha_{it}^k\varepsilon_{it}^k$ would also vanish as $n \to \infty$, at a rate $O\left(n^{1-\frac{k}{2}}\right)$, provided that $E\left|\varepsilon_{it}\right|^k < \infty$. In light of this, we propose to modify $CS_t^{*(k)}$ as

$$CS_t^{(k)} = n^{\frac{k}{2}-1}CS_t^{*(k)}. \tag{11}$$

It can be expected that, as $n \to \infty$, $CS_t^{(k)}$ converges to $(-1)^k \lim_{n\to\infty}\sum_{i=1}^{n}\alpha_{it}^k E\left(\varepsilon_{it}^k\right)$.

Let $\Gamma = \left[\gamma_2, ..., \gamma_p\right]'$ and $\widetilde{CS}_t = \left[CS_t^{(2)}, ..., CS_t^{(p)}\right]'$. The following Assumption, which complements Assumption 2, summarizes the discussion above.

Assumption 3: for all n it holds that *(i)* $E\left[\left(\alpha_t'\varepsilon_t\right)^p\right] < \infty$ and $0 < n^{\frac{p}{2}-1}\sum_{i=1}^{n}E\left[\left(\alpha_{it}^p\varepsilon_{it}^p\right)^2\right] < \infty$ and *(ii)* $E\left(\widetilde{CS}_t\widetilde{CS}_t'\right)$ is positive definite.

Based on the discussion above, equation (2) could be augmented as

$$\begin{aligned} y_t &= \beta y_{t-1} + \sum_{k=2}^{p}\gamma_k CS_t^{(k)} + v_t^* \\ &= \beta y_{t-1} + \Gamma'\widetilde{CS}_t + v_t^*. \end{aligned} \tag{12}$$

Similarly to Proposition 1, it holds that:

Theorem 1 *Let Assumptions 1-3 hold with* $E\left\|\varepsilon_t\right\|^{2p} < \infty$ *and consider*

$$\min_{\Gamma} E\left[v_t^*\left(\Gamma\right)\right]^2.$$

This has solution

$$\Gamma^* = \left[E\left(\widetilde{CS}_t\widetilde{CS}_t'\right)\right]^{-1}\left[E\left(\widetilde{CS}_t\alpha_t'\varepsilon_t\right)\right].$$

Also, it holds that $E\left[v_t^*\left(\Gamma^*\right)\right]^2 \leq E\left[v_t^2\right]$, *with* $E\left[v_t^*\left(\Gamma^*\right)\right]^2 = E\left[v_t^2\right]$ *if and only if* $\Gamma^* = 0$. *The same result holds as* $n \to \infty$, *assuming that* $\sup_i E\left|\varepsilon_{it}\right|^{2p} < \infty$ *and* $\sup_i \left|\alpha_{it}\right| = O\left(n^{-1/4}\right)$.

Theorem 1 is similar to Proposition 1. Particularly, gains are present if at least one element of Γ^* is non-zero, i.e. if at least one element of $E\left(\widetilde{CS}_t\alpha_t'\varepsilon_t\right)$ is non-zero, viz.

$$\sum_{j=0}^{k}\frac{1}{n^{k-j}}\binom{k}{j}E\left[\left(\alpha_t'\varepsilon_t\right)^{k-j}\sum_{i=1}^{n}\alpha_{it}^j\varepsilon_{it}^j\right] \neq 0,$$

for some k. Of course, in order to use this approach one needs the further assumption that $E\left\|\varepsilon_t\right\|^{2p} < \infty$. However, in this case, the MSE could be further reduced. Similarly to Proposition 1, in general the i-th forecaster is not able to use the true values of β and α_i. From the researcher's point of view, this entails that the $CS_t^{(k)}$s are computed based on possibly inconsistent estimates of the α_is. Even in this case, this is not necessarily a problem for the researcher, and using the $CS_t^{(k)}$s can still help improve the prediction of y_t.

Another advantage of using the $CS_t^{(k)}$s is that the dimensionality of the estimation problem is reduced. Indeed, if one were to estimate the α_{it}s in equation (1), even assuming homogeneity over time (i.e. $\alpha_{it} = \alpha_i$ for all t), as $n \to \infty$ this would be a classical incidental parameters problem (Neyman and Scott, 1948). Conversely, consider the OLS estimator of Γ^*, say $\hat{\Gamma}$, in the regression $y_t = \beta y_{t-1} + \Gamma'\widetilde{CS}_t + v_t^*$. Under the martingale difference assumption for the ε_{it}s, $T^{-1}\sum_t \widetilde{CS}_t y_{t-1} = O_p\left(T^{-1/2}\right)$ and thus $\hat{\Gamma} = \Gamma^* + O_p\left(T^{-1/2}\right)$.

Finally, it is interesting to note that the is need not represent different individuals. As a possible, alternative example, the equation $y_t^i = \beta y_{t-1} + \alpha_{it}\varepsilon_{it}$ could be the prediction generated from model i (which augments the AR(1) model by using a set of regressors ε_{it}) out of n possible models. In this case, Proposition 1 and Theorem 1 provide guidelines as to how to combine forecasts, as well as (see remarks above) spelling out the distributional assumptions on the regressors ε_{it} that make the combined forecast better than the basic AR(1) model.

The results in Proposition 1 and in Theorem 1 illustrate how measures of forecast disagreement can help to improve the forecast of y_t, especially under the realistic case of presence of cross dependence.

3 Applications

In this section, we first present a Monte Carlo exercise, to assess the impact of cross dependence (and other distributional properties) of the ε_{it}s on the ability of the augmented model (6) to yield better forecasts for y_t, using synthetic data. Secondly, we validate our findings with an application to US NFP data.

3.1 Monte Carlo simulations

The design of our experiments is as follows. We generate $T + 1000$ datapoints (discarding the first 1000 to avoid dependence on initial conditions) for y_t using equation (2). The alternative sample sizes we use are $T \in \{50, 100, 200\}$. Also, we set β in (2) equal to 0.5; this value is chosen based on the actual first order partial autocorrelation in the dataset used in Section 3.2. Other unreported results show that changing β has virtually no impact on the results. As far as the number of forecasters is concerned, we set $n \in \{15, 20, 25, 30, 45, 60, 80\}$.

We generate η_t as i.i.d. normal with zero mean and variance $\sigma_\eta^2 = 1$ (this parameter, too, does not appear to have much impact). We draw the α_{it}s from $iidN\left(n^{-1/2}, 1\right)$, so that Assumption 2 is satisfied; cross dependence among the ε_{it}s is modelled by setting, for $i \neq j$, $E\left(\varepsilon_{it}\varepsilon_{jt}\right) = \omega \in \{0, 0.2, 0.4, 0.6, 0.8, 0.99\}$.

The impact of asymmetry and cross dependence in the distribution of the ε_{it}s is analysed by generating them as (centered and scaled) chi-squared with p degrees of freedom. Experiments are carried out with $p = \{1, 5, 10, 30, 50\}$. As a benchmark, we also report an experiment where $\varepsilon_{it} \sim N(0, 1)$ - according to the theory, there should be no gain at all in this case when using $CS_t^{(2)}$. We also consider using the $CS_t^{(k)}$s when $\varepsilon_{it} \sim N(0, 1)$. In particular, we proceed in the following way. We include $CS_t^{(2)}$ and $CS_t^{(3)}$ only for $T = 50$, in order not to saturate the degree of freedom. When $T = 100$, we add $CS_t^{(5)}$ and $CS_t^{(6)}$ (as well as $CS_t^{(2)}$ and $CS_t^{(3)}$); finally, we consider $CS_t^{(2)}$ up to $CS_t^{(8)}$ in the case $T = 200$. In these cases, Theorem 1 states that there should be some gains in predictive ability.

We measure the gain in predictive ability by using in-sample forecasts for all $t = 1, ..., T$. Let MSE_1 and MSE_2 be the Mean Squared Errors from models (2) and (12) respectively. The values in Table 1 are calculated as

$$gain = -\frac{MSE_2 - MSE_1}{MSE_1}. \tag{13}$$

The number of replications is $10,000$.

[Insert Table 1 somewhere here]

The results complement Proposition 1 and Theorem 1. As p increases, the distribution of the ε_{it}s approaches a Gaussian distribution; as a consequence, gains become increasingly smaller. Also, gains monotonically decrease as ω, the degree of cross-dependence, decreases.

As predicted by the theory, in the case of Gaussian ε_{it}, there are no improvements in predictive power when including $CS_t^{(2)}$ in (6). However, as the last two panels of the table show, including the $CS_t^{(k)}$s seems to yield some reduction in the MSE, in contrast to the case of Gaussian signals with $CS_t^{(2)}$ as a proxy; larger sample sizes, which allow for higher order $CS_t^{(k)}$s, show a moderate improvement in predictive ability. It is interesting to explore the link between the chi-squared and the Gaussian case. When p is as large as 30, and there is no cross dependence ($\omega = 0$), there is no gain from adding $CS_t^{(2)}$. This follows from the theory: as $p \to \infty$, the Central Limit Theorem entails that the distribution of ε_{it} is tantamount to a normal distribution. In this case, predictive ability is present when there is a large amount of cross dependence. This is probably due to the fact that, when ω is large, the convergence to the normal distribution gets slower. Table 1 also shows the role played by the number of forecasters n: irrespective of the distributional properties of the private signal ε_{it}, increases in n amplify the results, and particularly the spread between MSE gains when $\omega = 0$ as opposed to $\omega = 0.99$.

3.2 Empirical exercise

In order to validate the use of the $CS_t^{(k)}$s studied in Proposition 1 and Theorem 1, we report an illustrative application based on predicting a "classical" economic indicator, namely (changes in the) US NFP data (y_t). Our monthly dataset spans from June 2000 until July 2004 (thus, $T = 50$); the number of forecasters, n_t, increases over time, ranging between 37 and 79 with a median value of 56.

We calculate $CS_t^{(k)}$ as defined in (11), considering $k = 2, 3$ and 4. Descriptive statistics for all the series are reported in Table 2, where we also report the correlogram for y_t. Preliminary analysis shows that $CS_t^{(4)}$ is non-stationary; thus, we use its first difference, $\Delta CS_t^{(4)}$, whose descriptive statistics are reported in the Table. The correlogram of y_t shows a quite clear AR(1) pattern.

[Insert Table 2 somewhere here]

We now turn to analysing the output. We compare the predictive ability of four different models:

$$
\begin{aligned}
&\text{Model 1:} \quad && y_t = \alpha + \beta y_{t-1} + v_t, \\
&\text{Model 2:} \quad && y_t = \alpha + \beta y_{t-1} + \gamma_1 \bar{y}_t + v_t^*, \\
&\text{Model 3:} \quad && y_t = \alpha + \beta y_{t-1} + \gamma_1 \bar{y}_t + \gamma_2 CS_t^{(2)} + v_t^*, \\
&\text{Model 4:} \quad && y_t = \alpha + \beta y_{t-1} + \gamma_1 \bar{y}_t + \gamma_2 \bar{y}_{t-1} + \gamma_3 CS_t^{(3)} + \gamma_4 CS_{t-1}^{(3)} + v_t^*,
\end{aligned}
$$

where, as above, $\bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_t^i$, i.e. it is the mean of the individual forecasts. This is the most obvious proxy for private information, and the purpose of the exercise is to verify whether augmenting the model with the $CS_t^{(k)}$s can significantly enhance forecasting ability. Model 1 is used as a benchmark, and it is a standard AR(1) model as identified using the correlogram in Table 2. Model 2 augments the baseline AR(1) specification by using $\bar{y}_t$ as a proxy for the unobservable private information. Building on Proposition 1 and Theorem 1, we preliminarily consider Model 3, which is based on augmenting Model 2 by using the "traditional" measure of disagreement $CS_t^{(2)}$. As we discuss later on in greater detail, $CS_t^{(2)}$ is found to be irrelevant. Thus, as mentioned in the Introduction, we take a "non structural" approach to modelling y_t. Indeed, this is our recommended approach: the $CS_t^{(k)}$s do not have a structural interpretation, and it is possible that y_t may also depend on past values of the $CS_t^{(k)}$s, due to the possible presence of inertia in the individual forecasters' predictions. Thus, we suggest a GETS approach, by estimating, as a Generalised Unrestricted Model (GUM), the following ADL model for y_t:

$$\begin{aligned} y_t &= \alpha + \beta y_{t-1} + \gamma_{1,0}\bar{y}_t + \gamma_{1,1}\bar{y}_{t-1} + \gamma_{2,0}CS_t^{(2)} + \gamma_{2,1}CS_{t-1}^{(2)} + \\ &\quad \gamma_{3,0}CS_t^{(3)} + \gamma_{3,1}CS_{t-1}^{(3)} + \gamma_{4,0}\Delta CS_t^{(4)} + \gamma_{4,1}\Delta CS_{t-1}^{(4)} + v_t^*. \end{aligned} \tag{14}$$

Preliminary analysis carried out using the AutoMetrics option in OxMetrics 6.2 shows that relevant explanatory variables in the model are, in addition to $\bar{y}_t$ and $\bar{y}_{t-1}$, also $CS_t^{(3)}$ and $CS_{t-1}^{(3)}$, whence Model 4.

The goodness of fit of each model is assessed using the adjusted R^2, computed for the whole sample $t = 1, ..., T$. As far as forecasting ability is concerned, comparisons are based on the MSE. Note that Models 2, 3 and 4 all nest Model 1 (see e.g. Clark and McCracken, 2001, 2005, 2006; Clark and West, 2007). We construct the predictions for y_t using a recursive scheme (West, 2005; Clark and West, 2007); Model 4 also nests Model 2. This entails firstly estimating the models using data from $t = 1$ up to $t = R$, and use the estimated parameters to predict y_{t+R+1}; the estimates are then recalculated using all available data from $t = 1$ up to $t = R + 1$, and the prediction of y_{t+R+2} is calculated, and so on[1]. In our context, we carry out predictions from July 2002 (at mid-sample) until the end of the sample, so that $R = 25$ and the number of predictions is $P = 25$.

Let MSE_i be the Mean Squared Error associated with model i. We carry out the relevant tests based on the following framework

$$\begin{cases} H_0 : MSE_i = MSE_j \\ H_0 : MSE_i < MSE_j \end{cases} \text{for } i \neq j.$$

Tests are based on using the adjusted MSE statistic discussed in Clark and West (2007). Letting $\hat{e}_{i,t+1}$ be the forecast error for y_{t+1} made by Model i, the adjusted MSE is defined as

$$MSE_{ij}^{adj} = \frac{2}{P} \sum_{t=R+1}^{T} \hat{e}_{j,t+1} \left(\hat{e}_{j,t+1} - \hat{e}_{i,t+1} \right) = \frac{1}{P} \sum_{t=R+1}^{T} \omega_{ij,t}, \tag{15}$$

[1] Other schemes are possible, e.g. the "rolling" one, where the estimation sample has the same size, R, so that y_{t+R+1} is predicted using estimates using the sample $t = 1, ..., R$; y_{t+R+2} using estimates from the sample $t = 2, ..., R + 1$, and so on.

using the compact notation $\omega_{ij,t} = 2\hat{e}_{j,t+1}\ (\hat{e}_{j,t+1} - \hat{e}_{i,t+1})$. The variance of $\omega_{ij,t}$ is estimated by a HAC-type estimator (we define the estimate as $\hat{\sigma}^2_{\omega ij})^2$. The test statistic that we use, t^{enc}_{ij}, is discussed by Clark and McCracken (2001), and it is defined as

$$t^{enc}_{ij} = \frac{1}{\hat{\sigma}^2_{\omega ij}\sqrt{P}} \sum_{t=R+1}^{T} \omega_{ij,t}. \tag{16}$$

One attractive computational feature of t^{enc}_{ij} is that, although t^{enc}_{ij} does not follow the standard normal distribution as $R, P \to \infty$, using quantiles from the standard normal yields mildly conservative tests - e.g. Clark and West (2007, p. 298-299) argue that using 1.645 as a critical value yields a test of size between 0.01 and 0.05 for R and P large enough. Thus, we base our tests on standard normal inference, as indicated by Clark and West (2007).

Results (alongside with estimation output and mis-specification tests) are reported in Table 3:

[Insert Table 3 somewhere here]

Consider Model 2. The estimation output shows that y_{t-1} is not significant, whilst $\bar{y}_t$ is significant. The $\bar{R}^2$ increases by around 0.25 compared to that of Model 1; as far as predictive ability is concerned, we note that the MSE decreases by around 30% with respect to Model 1. Moreover, a test based on t^{enc}_{21} shows that $MSE_2 < MSE_1$. Turning to Model 3, the output clearly shows that $CS^{(2)}_t$ is not significant. This is reinforced by the fact that the MSE is virtually unchanged from Model 2. As pointed out in the comment to Proposition 1, this may be due to a plurality of reasons (e.g. the ε_{it}s being Gaussian), but the results show that there is no gain in predictive ability - we did not carry out a test for $H_0 : MSE_3 = MSE_2$ as the outcome is already quite clear.

Finally, consider the recommended modelling strategy, Model 4. From the GUM (14), we obtained a final model containing $\bar{y}_t$ and $CS^{(3)}_t$ and their first lags. Inspecting the significance of parameters, y_{t-1} is barely significant (at a 10% level); both $\bar{y}_t$ and $\bar{y}_{t-1}$ are significant, which is partly in line with Model 2; and, finally, $CS^{(3)}_{t-1}$ is significant, whereas $CS^{(3)}_t$ is borderline significant. Model 4 has superior explanatory power with respect to Model 1: the $\bar{R}^2$ increases by more than double. Turning to forecasting ability, the MSE declines sharply, by 50%. Further, this is a significant decline, in view of t^{enc}_{41}: the null that $MSE_4 = MSE_1$ is rejected at the 5% level. Also, Model 4 is shown to be better than Model 2: by using t^{enc}_{42}, the null that $MSE_4 = MSE_2$ is rejected at the 5% level. Indeed, as we point out above, standard normal inference using t^{enc}_{42} tends to be mildly conservative (Clark and West, 2007), which reinforces the

[2] We compute $\hat{\sigma}^2_{\omega i}$ based on Andrews (1991). Specifically, we use a Bartlett kernel. Data are pre-whitened by fitting an AR model whose order is selected using Akaike Information Criterion; see Andrews and Monahan (1992).

conclusion that $MSE_4 < MSE_2$. Thus, it can be concluded from this example that the $CS_t^{(k)}$ can add significant predictive power on top of the mean forecast $\bar{y}_t$.

4 Concluding remarks

The main aim of this paper was to study how to extract private information from individual forecasts, and how to use such private information in order to enhance the predictive ability for an outcome variable y_t. We define a class of measures of cross sectional dispersion (defined as $CS_t^{(k)}$) which are related to the sample moments of the cross section of forecasts. We find that, in presence of cross sectional dependence, such measures are useful to increase forecasting accuracy for y_t, by proxying private information. The theory developed in Section 2 clearly shows that the usefulness of the $CS_t^{(k)}$s depends on the presence and amount of cross dependence across forecasts, which is a well documented fact in empirical applications.

From a methodological point of view, the results in the empirical part of this paper suggest some guidelines on how to use the $CS_t^{(k)}$s. In view of the non structural nature (and in view of the lack of a structural interpretation for them), we recommend employing a GETS approach, by starting, as a GUM, from an ADL specification, thereby using lags of y_t and of the measures of cross sectional dispersion, $CS_t^{(k)}$. These findings are reinforced through an application to the US NFP data. Of course, results in Section 3.2 refer to one dataset only, however important, and in order to validate the theory developed here it is necessary to undertake a substantive set of empirical applications.

Acknowledgment

This paper was previously circulated under the working title "On the Relationship between Cross Sectional and Time Series Measures of Uncertainty". We thanks the Editor (Michael Clements), an Associate Editor and three anonymous Referees for very constructive feedback. We are very grateful to participants in the The Royal Economic Society Conference (Swansea 7-9, April 2004), the North American Summer Meeting of the Econometrics Society (Providence 17-20, June 2004), in particular John Sutton, the ESRC Econometric Study Group Conference (Bristol, 14-16 July, 2005), in particular Peter Boswijk and David Hendry, the IFO Conference on "Survey Data in Economics - Methodology and Applications", with special mention to our discussant Klaus Abberger, and North American Summer Meeting of the Econometric Society (Duke, 21-24 June, 2007), in particular Valentina Corradi and Peter R. Hansen, for useful suggestions and comments. Special thanks to Roy Batchelor and Peter Pope for detailed comments on a previous version of the paper. We gratefully acknowledge Daniel Braberman's help with providing the data. The usual disclaimer applies. ESRC funding under grant R000223385 is gratefully acknowledged. L. Trapani acknowledges financial support from Cass Business School under the RAE Development Fund scheme and ESRC Postdoctoral Fellowship Scheme (PTA-026-27-1107).

References

Andrews, D.W.K (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817-858.

Andrews, D.W.K, Monahan, J.C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60, 953-966.

Barron, O.E., Kim, O., Lim, C.S., Stevens, D.E. (1998). Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review*, 73, 421-433.

Castle, J.L., Doornik, J.A., and Hendry, D.F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*, 3(1), Article 8.

Clark, T.E., McCracken, M.W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105, 85–110.

Clark, T.E., McCracken, M.W. (2005). Evaluating direct multistep forecasts. *Econometric Reviews*, 24, 369–404.

Clark, T.E., McCracken, M.W. (2006). The predictive content of the output gap for inflation: resolving in-sample and out-of-sample evidence. *Journal of Money, Credit, and Banking*, 38, 1127-1148.

Clark, T.E., West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138, 291-311.

Chamley, C.P. (2004). Rational herds: economic models of social learning. Cambridge University Press: Cambridge.

Doornik, J. (2009). Autometrics. In: Castle, J. and Shpehard, N., The Methodology and Practice of Econometrics, 1(9), 88–122. Oxford: Oxford University Press.

Dovern, J., Fritsche, U., Slacalek, J. (2011). Disagreement among forecasters in G7 countries. Forthcoming, *The Review of Economics and Statistics*.

Elliott, G. (2011). Averaging and the optimal combination of forecasts. Mimeo, University of California San Diego.

Engle, R.F. (1983). Estimates of the variance of U.S. inflation based upon the ARCH model. *Journal of Money, Credit and Banking*, 15, 286-301.

Fischer, P.E., Verrecchia, R.E. (1998). Correlated forecast errors. *Journal of Accounting Research*, 36, 91-110.

Genre, V., Kenny, G, Meyler, A., Timmermann, A. (2010). Combining the forecasts in the ECB survey of professional forecasters: can anything beat the simple average? European Central Bank Working Paper Series 1277.

Giordani, P., Soderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, 47, 1037-1059.

Gregory, A.W., Smith, G.W., Yetman, J. (2001). Testing for forecast consensus. *Journal of Business and Economic Statistics*, 19, 34-43.

Lahiri, K., Sheng, X. (2010). Measuring forecast uncertainty by disagreement: the missing link. *Journal of Applied Econometrics,* 25, 514-538.

Lys, T., Sohn, S. (1990). The association between revisions of analysts' earnings forecasts and security-price changes. *Journal of Accounting and Economics*, 341-63.

Neyman, J., Scott, E.L. (1948). Consistent estimation from partially consistent observations. *Econometrica*, 16, 1-32.

O'Brien, P. (1988). Analysts' forecasts as earnings expectations. *Journal of Accounting and Economics*, 53-83.

Patton, A.J., Timmermann, A. (2010). Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, 57, 803-820.

Pesaran, M.H., Weale, M.R. (2006). Survey expectations. In: *Handbook of Economic Forecasting* (eds G. Elliott, C.W.J. Granger and A. Timmermann). Elsevier, North-Holland.

Scharfstein, D.S., Stein, J.C. (1990). Herd behavior and investment. *American Economic Review*, 80, 465-79.

Stein, J.C. (2003). Agency, information and corporate investment. In: *Handbook of the Economics of Finance* (eds. G.M. Constantinides, M. Harris, and R.M. Stulz). Elsevier, North-Holland.

West, K.D. (2005). Forecast evaluation. University of Wisconsin (manuscript).

Ciaran Driver (PhD, CNAA) is currently Professor of Economics in the Department of Finance and Management Studies at the School of Oriental and African Studies, University of London. His main research interests are on capital investment at different level of aggregation in terms of asset type, country and sector, and on productivity spillovers and regional innovation. Ciaran is the author or editor of 4 books and of several academic articles and notes in the *Journal of Business and Economic Statistics*, *Economic Journal*, *European Economic Review*, *American Economic Review*, *Economics Letters*, *Journal of Economic Behavior and Organization*, *Oxford Bulletin of Economics and Statistics*, *International Journal of Industrial Organization* and others.

Lorenzo Trapani (PhD, University of Bergamo) is currently Senior Lecturer in Finance at Cass Business School, London. His main research interests are in econometric theory, particularly asymptotic theory, panel data, rank tests and testing for structural breaks. Recent publications include *Econometric Reviews*, *Econometric Theory*, *International Journal of Forecasting* and *Journal of Econometrics*.

Giovanni Urga (PhD, Oxford) is Professor of Finance and Econometrics and Director of Centre for Econometric Analysis (CEA@Cass) at Cass Business School, London. His research interests are panel data, financial econometrics, modelling risk and cross-market correlations, asset pricing, structural breaks, modelling common stochastic trends, credit spreads. Recent publications include the *Journal of Econometrics*, *Journal of Business and Economic Statistics*, *Economics Letters*, *Econometric Theory*, *Oxford Bulletin of Economics and Statistics*, *Journal of Applied Econometrics* and others. He is Associate Editor for *Empirical Economics* and he has

been guest editor for the *Journal of Econometrics* and the *Journal of Business and Economic Statistics*.

					$p = 1$			
	n	15	20	25	30	45	60	80
ω	T							
	50	7.67	7.14	6.53	5.65	4.10	4.13	2.66
0	100	7.41	6.35	6.16	5.30	4.01	3.31	2.90
	200	7.27	6.86	5.71	5.27	3.92	3.19	2.44
	50	11.93	12.56	13.03	12.86	12.80	13.94	14.35
0.2	100	10.94	11.51	12.54	12.73	13.24	13.26	14.02
	200	11.35	11.94	11.79	12.25	12.78	13.12	13.93
	50	20.03	23.25	24.89	25.86	28.06	30.04	32.40
0.4	100	19.12	22.13	24.48	25.67	28.69	30.20	32.00
	200	20.36	22.21	23.95	25.56	28.32	29.96	31.77
	50	28.37	32.98	35.69	36.83	40.13	42.31	44.73
0.6	100	27.84	32.29	34.73	36.63	40.54	42.97	44.85
	200	29.34	32.08	34.79	36.99	40.27	42.51	44.44
	50	35.79	40.63	43.69	44.98	48.62	50.70	52.74
0.8	100	35.27	40.36	42.27	44.68	48.62	51.67	53.19
	200	36.58	39.98	42.81	45.20	48.49	51.10	52.80
	50	41.71	46.56	48.66	50.58	54.24	56.21	58.15
0.99	100	41.38	46.23	47.55	49.95	53.86	57.19	58.33
	200	41.87	45.82	48.27	50.87	54.29	57.08	58.28

					$p = 5$			
		15	20	25	30	45	60	80
ω	T							
	50	2.78	2.45	2.09	1.78	1.42	1.32	0.80
0	100	2.64	1.96	1.88	1.67	1.17	0.74	0.82
	200	2.55	2.08	1.76	1.54	0.98	0.80	0.56
	50	4.58	4.36	4.55	4.16	4.19	4.32	4.29
0.2	100	4.30	3.96	3.93	4.08	4.15	3.63	3.97
	200	4.11	4.05	3.92	3.82	3.57	3.68	3.60
	50	8.27	9.03	9.70	9.50	10.21	10.53	10.94
0.4	100	8.00	8.48	8.76	9.14	10.14	9.92	10.70
	200	7.83	8.54	9.04	9.01	9.45	10.03	10.09
	50	12.34	13.96	14.85	15.07	16.19	16.56	17.10
0.6	100	12.14	13.33	13.87	14.38	15.90	15.97	17.07
	200	12.11	13.41	14.52	14.55	15.37	16.26	16.46
	50	16.10	18.23	19.34	19.97	21.11	21.60	22.14
0.8	100	15.90	17.64	18.29	19.04	20.75	21.02	22.22
	200	16.10	17.75	19.29	19.48	20.37	21.53	21.82
	50	19.23	21.88	23.08	24.18	24.51	25.49	25.99
0.99	100	18.98	21.02	21.71	22.96	24.66	25.02	25.93
	200	19.57	21.19	23.00	23.54	24.36	25.80	25.95

	n	15	20	25	30	45	60	80
					$p = 10$			
ω	T							
0	50	1.63	1.41	1.30	1.11	0.83	0.81	0.56
	100	1.52	1.03	1.10	0.97	0.59	0.48	0.48
	200	1.42	1.12	0.90	0.79	0.50	0.43	0.32
0.2	50	2.66	2.39	2.50	2.45	2.38	2.35	2.17
	100	2.43	2.19	2.28	2.29	2.13	1.88	2.21
	200	2.30	2.27	2.01	2.04	1.86	1.96	1.97
0.4	50	4.84	5.04	5.30	5.55	5.86	5.79	5.70
	100	4.42	4.82	5.03	5.14	5.51	5.40	6.01
	200	4.51	4.91	4.88	5.00	5.12	5.45	5.66
0.6	50	7.28	7.88	8.54	8.82	9.37	9.26	9.36
	100	6.76	7.81	8.04	8.27	8.97	9.09	9.82
	200	7.07	7.85	8.16	8.25	8.57	9.12	9.43
0.8	50	9.52	10.33	11.70	11.84	12.31	12.29	12.63
	100	9.03	10.67	10.75	11.13	12.00	12.32	13.09
	200	9.44	10.50	11.12	11.22	11.67	12.44	12.73
0.99	50	11.50	12.32	14.30	14.67	14.32	14.66	15.21
	100	11.07	13.02	12.98	13.57	14.53	14.88	15.59
	200	11.48	12.48	13.41	13.84	14.32	15.22	15.34

		15	20	25	30	45	60	80
					$p = 30$			
ω	T							
0	50	0.70	0.71	0.73	0.68	0.51	0.51	0.32
	100	0.58	0.39	0.52	0.44	0.26	0.22	0.23
	200	0.50	0.42	0.37	0.31	0.21	0.16	0.10
0.2	50	1.08	1.00	1.15	1.19	1.08	1.05	0.96
	100	0.90	0.77	0.88	1.00	0.87	0.69	0.91
	200	0.83	0.79	0.75	0.75	0.66	0.70	0.63
0.4	50	1.91	1.96	2.29	2.33	2.37	2.20	2.30
	100	1.66	1.64	1.71	2.07	2.15	1.93	2.27
	200	1.68	1.70	1.78	1.84	1.86	1.95	1.91
0.6	50	2.87	3.07	3.62	3.64	3.74	3.46	3.64
	100	2.58	2.72	2.74	3.24	3.40	3.37	3.67
	200	2.71	2.75	3.04	3.09	3.21	3.32	3.31
0.8	50	3.80	4.02	4.87	4.92	5.04	4.68	4.77
	100	3.52	3.85	3.77	4.39	4.51	4.66	4.90
	200	3.71	3.79	4.28	4.29	4.47	4.58	4.61
0.99	50	4.72	4.68	5.91	6.16	6.20	5.72	5.54
	100	4.41	4.86	4.67	5.42	5.49	5.60	5.86
	200	4.63	4.84	5.28	5.35	5.55	5.61	5.72

	n	15	20	25	30	45	60	80
					$p = 50$			
ω	T							
0	50	0.55	0.65	0.63	0.49	0.39	0.48	0.29
	100	0.35	0.25	0.33	0.28	0.16	0.14	0.17
	200	0.33	0.27	0.25	0.20	0.16	0.11	0.07
0.2	50	0.77	0.81	0.84	0.91	0.70	0.80	0.54
	100	0.57	0.50	0.56	0.64	0.52	0.43	0.59
	200	0.54	0.49	0.49	0.47	0.43	0.45	0.39
0.4	50	1.27	1.39	1.49	1.72	1.35	1.52	1.46
	100	1.04	1.06	1.03	1.27	1.31	1.10	1.40
	200	1.07	1.01	1.13	1.16	1.15	1.18	1.18
0.6	50	1.89	1.97	2.36	2.52	2.10	2.36	2.39
	100	1.60	1.77	1.63	1.94	2.12	1.91	2.05
	200	1.71	1.61	1.92	1.93	1.99	1.99	2.05
0.8	50	2.48	2.49	3.21	3.28	2.85	3.18	3.22
	100	2.11	2.48	2.29	2.59	2.82	2.72	3.04
	200	2.34	2.22	2.68	2.67	2.76	2.77	2.86
0.99	50	3.01	2.95	3.80	3.91	3.50	3.91	3.84
	100	2.53	3.02	2.84	3.28	3.38	3.47	3.67
	200	2.85	2.87	3.29	3.33	3.40	3.40	3.52

					$N(0,1)$			
		15	20	25	30	45	60	80
ω	T							
0	50	0.30	0.24	0.22	0.18	0.19	0.22	0.19
	100	0.09	0.08	0.06	0.07	0.07	0.05	0.05
	200	0.04	0.03	0.04	0.03	0.02	0.02	0.02
0.2	50	0.31	0.25	0.21	0.21	0.22	0.22	0.20
	100	0.08	0.06	0.07	0.07	0.06	0.06	0.06
	200	0.04	0.03	0.03	0.03	0.02	0.02	0.01
0.4	50	0.31	0.24	0.21	0.21	0.22	0.21	0.19
	100	0.08	0.06	0.07	0.07	0.06	0.06	0.06
	200	0.04	0.03	0.03	0.03	0.02	0.02	0.02
0.6	50	0.30	0.24	0.21	0.21	0.22	0.20	0.19
	100	0.08	0.06	0.06	0.08	0.06	0.05	0.06
	200	0.04	0.03	0.03	0.03	0.02	0.02	0.02
0.8	50	0.30	0.24	0.21	0.21	0.22	0.20	0.20
	100	0.08	0.06	0.08	0.08	0.06	0.05	0.06
	200	0.03	0.03	0.03	0.03	0.02	0.02	0.02
0.99	50	0.28	0.24	0.20	0.21	0.21	0.21	0.19
	100	0.09	0.07	0.08	0.08	0.05	0.05	0.06
	200	0.03	0.03	0.03	0.03	0.02	0.02	0.02

$N(0,1) + CS_t^{(3),(4),(5),(6),(7),(8)}$

ω	n / T	15	20	25	30	45	60	80
0	50	1.71	1.52	1.48	1.52	1.54	1.42	1.35
	100	1.70	1.72	1.62	1.69	1.59	1.66	1.62
	200	1.30	1.35	1.34	1.29	1.27	1.30	1.26
0.2	50	1.71	1.60	1.47	1.60	1.54	1.47	1.29
	100	1.69	1.70	1.65	1.72	1.68	1.63	1.64
	200	1.30	1.33	1.35	1.29	1.31	1.28	1.29
0.4	50	1.69	1.61	1.47	1.59	1.53	1.48	1.31
	100	1.67	1.68	1.71	1.73	1.71	1.63	1.65
	200	1.31	1.33	1.35	1.29	1.30	1.29	1.30
0.6	50	1.68	1.61	1.48	1.58	1.52	1.47	1.33
	100	1.67	1.66	1.74	1.73	1.73	1.62	1.65
	200	1.30	1.32	1.36	1.28	1.29	1.29	1.31
0.8	50	1.67	1.62	1.48	1.57	1.50	1.46	1.34
	100	1.68	1.66	1.74	1.71	1.74	1.62	1.64
	200	1.31	1.31	1.36	1.28	1.30	1.29	1.31
0.99	50	1.66	1.64	1.48	1.58	1.50	1.44	1.36
	100	1.67	1.67	1.70	1.68	1.71	1.65	1.66
	200	1.30	1.30	1.34	1.29	1.29	1.30	1.30

Table 1. The values in the table are MSE gains, as defined in (13). In each table, the first column contains the degree of cross dependence, ω. Tables whose headings contain p indicate the degree of freedom of the chi-squared distributions used to generate ε_{it}; tables whose headings are $N(0,1)$ and $N(0,1) + CS_t^{(3),(4),(5),(6),(7),(8)}$ refer to the cases where ε_{it} follows a standard normal and equation (2) is augmented using $CS_t^{(2)}$ only and the $CS_t^{(k)}$s respectively. For the latter, we refer to the main text.

Descriptive statistics					Correlogram of y_t		
	y_t	$CS_t^{(2)}$	$CS_t^{(3)}$	$\Delta CS_t^{(4)}$	Lag	ACF	PACF
Mean	0.038	18.227	58.508	165.840	1	0.543^{**}	0.543^{**}
Median	0.035	13.307	80.898	71.923	2	0.324^{**}	0.042
Max.	3.080	66.638	469.01	25697.87	3	0.172	−0.027
Min.	−4.150	4.353	−960.609	−26873.68	4	0.186	0.131
Std. Dev.	1.534	13.155	232.700	6445.930	5	0.094	−0.075
					6	0.029	−0.041
Bera-Jarque	0.984	37.680^{***}	84.386^{***}	163.813^{***}	7	−0.053	−0.066
					8	−0.070	−0.029
ADF	-3.913^{***}	-4.521^{***}	-5.223^{***}	-10.209^{***}	9	−0.193	−0.183
					10	−0.246	−0.093
					11	−0.257	−0.046
					12	−0.203	−0.016
					13	−0.216	−0.071

Table 2. Descriptive statistics and correlogram of y_t (the latter contains, respectively, autocorrelations, ACF, and partial autocorrelations, PACF). For the Bera-Jarque and the Augmented Dickey-Fuller (ADF) tests, the value of the test statistics has been reported; the symbol (***) indicates rejection at 1% level. The symbol (**) in the correlogram panel denotes rejection at 5% level of the null that the estimated autocorrelation or partial autocorrelation is zero.

	Model 1	Model 2	Model 3	Model 4
Estimation output				
y_{t-1}	0.560^{**}	-0.041	-0.103	0.261^{*}
	(0.112)	(0.143)	(0.144)	(0.145)
$\bar{y}_t$		1.07^{**}	1.12^{**}	1.290^{**}
		(2.00)	(0.198)	(0.226)
$\bar{y}_{t-1}$				-0.621^{**}
				(0.208)
$CS_t^{(2)}$			-1.930^{*}	
			(1.100)	
$CS_t^{(3)}$				-0.0012^{**}
				(0.00057)
$CS_{t-1}^{(3)}$				0.0017^{**}
				(0.00060)
Mis-Spec. Tests				
AR 1-4	0.3642	1.7824	2.738	2.509
	[0.833]	[0.150]	[0.041]	[0.057]
Heterosk.	0.0701	0.7071	0.3906	0.5199
	[0.932]	[0.591]	[0.881]	[0.866]
Ramsey's Reset	0.5561	2.9417	1.769	1.351
	[0.577]	[0.093]	[0.190]	[0.252]
Bera-Jarque	7.084	3.653	3.356	1.483
	[0.029]	[0.161]	[0.187]	[0.477]
Quandt-Andrews	4.883	5.749	5.757	6.098
	[0.578]	[0.698]	[0.867]	[0.414]
Goodness of fit				
$\bar{R}^2$	0.327	0.575	0.593	0.673
MSE	1.388	0.978	0.986	0.662
t_{ij}^{enc}	–	$t_{21}^{enc} = -1.7136^{**}$	–	$t_{41}^{enc} = -1.7409^{**}$
				$t_{42}^{enc} = -1.7880^{**}$

Table 3. Regression outputs for Models 1-4. Numbers in round brackets in the "Estimation output" section indicate standard errors; the symbol $(^{*})$ and $(^{**})$ denote rejection at 10% and 5% level respectively of the null that the corresponding coefficient is non significant. In the "Mis-specification tests" section, we report: the Breusch-Godfrey test carried out up to lag 4 (AR1-4); White's test for heteroskedasticity using squares only (Heterosk.); Ramsey's RESET test using only the square of the fitted value; the Bera-Jarque test for normality; Andrews' (1993) test for a break, reporting the Sup of the sequence of the Wald statistics, constructed by trimming the first and last 15% of the datapoints (Quandt-Andrews). Numbers in square brackets are the p-values. In the "Goodness of Fit" section of the Table, we report the adjusted R^2, the MSE for each model constructed as described in Section 3.2, and the t_{ij}^{enc} statistics defined in (16), for the null that Model i has the same forecasting accuracy as Model 1. We do not report the test statistic for Model 2 as the MSE is the same. The symbol $(^{**})$ denotes rejection at 5% level of the null.