ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

Nowcasting with Backcalculated Short Time Series Simulations & Empirical Evidence

George Kapetanios and Fotis Papailias

ESCoE Discussion Paper 2023-08

June 2023

ISSN 2515-4664

DISCUSSION PAPER

Nowcasting with Backcalculated Short Time Series Simulations & Empirical Evidence
George Kapetanios and Fotis Papailias
ESCoE Discussion Paper No. 2023-08
June 2023

Abstract

This paper is concerned with the backcalculation of short explanatory time series when the dependent variable is longer. In particular, we consider two competing linear regression models: (i) the main model which is estimated in the overlapping period when all time series are available, and (ii) the suggested model which produces the estimates in two steps; first, we create backcalculated values of the explanatory time series using some auxiliary variables (which are observed at the same -longer- time history as the target variable) in the overlapping time period and, using these coefficient estimates, we extend the short explanatory time series and estimate a model which regresses the dependent variable on this new set of variables. This research provides both simulations and empirical evidence in favour of the suggested method.

Keywords: Backcalculation, Model Selection

JEL classification: C32, C52

Fotis Papailias, King's Business School, Data Analytics Centre for Macro and Finance, Economic Statistics Centre of Excellence
fotis.papailias@kcl.ac.uk

Published by:
Economic Statistics Centre of Excellence
King's College London
Strand
London
WC2R 2LS
United Kingdom
www.escoe.ac.uk

Nowcasting with Backcalculated Short Time Series Simulations & Empirical Evidence*

George Kapetanios Fotis Papailias†
King's Business School
Data Analytics Centre for Macro and Finance
Economic Statistics Centre of Excellence

Abstract

This paper is concerned with the backcalculation of short explanatory time series when the dependent variable is longer. In particular, we consider two competing linear regression models: (i) the main model which is estimated in the overlapping period when all time series are available, and (ii) the suggested model which produces the estimates in two steps; first, we create backcalculated values of the explanatory time series using some auxiliary variables (which are observed at the same -longer- time history as the target variable) in the overlapping time period and, using these coefficient estimates, we extend the short explanatory time series and estimate a model which regresses the dependent variable on this new set of variables. This research provides both simulations and empirical evidence in favour of the suggested method.

Keywords: Backcalculation, Model Selection.
JEL Codes: C32, C52.

*We would like to thank our ONS colleagues, Bertrand Abang, Robert Kendall, Keith Lai, Silvia Lui, Sanjiv Mahajan, Craig McLaren, Ewan Moffat and Andrew Walton for insightful discussions and feedback which significantly improved our work. We also thank ESCoE's director, Rebecca Riley, for helpful comments at an earlier stage of this research, and ESCoE's Operations Director, Sarah Sheppard, for invaluable support. This research has been funded by the Office for National Statistics as part of the research programme of the Economic Statistics Centre of Excellence. Any remaining errors are ours.

† Corresponding author. King's Business School, King's College London, Bush House, 30 Aldwych, London, WC2B 4BG, UK. Knot Analytics Ltd. E-mail: fotis.papailias@kcl.ac.uk

Contents

1 Introduction

The advancements of technology during the past twenty years have given rise to the, now popular, field of big data. A vast amount of data can be stored, managed and efficiently analysed using the necessary CPU or GPU power. The field of economics is no different as researchers have been investigating a number of alternative sources and variables to supplement existing traditional economics and financial datasets.[1] The Office for National Statistics (ONS) in the UK has been supplying researchers with such "real-time" indicators since 2018.[2]

The recent COVID-19 pandemic further highlighted the need of real-time (possibly big data or big data-based) indicators and accelerated the publishing of such datasets. To satisfy the researchers' demand, and also to provide as many tools as possible to study the multi-dimensional effects of the pandemic, many organisations started publishing daily or weekly datasets since the late 2019 or early 2020. In particular, the ONS made available, among others, the following set of indicators: online job advertisements (weekly, February 2018), company incorporations and voluntary dissolutions (weekly, December 2018), advanced notification of potential redundancies (weekly, April 2019), shipping indicators (weekly, April 2019), spending on credit and debit cards (daily, January 2020), coronavirus and the social impacts on Great Britain (daily, March 2020), traffic camera activity (daily, March 2020) and coronavirus infection survey (daily, April 2020).

There is no doubt that the applied researcher needs as many datasets available as possible, meaning that the above-mentioned real-time indicators should be seriously considered. In fact, studies in the literature have already shown the potential benefits of the real-time information of these datasets in the production of flash estimates, nowcasting and in the construction of macroeconomic coincident indicators; see, e.g., Kapetanios and Papailias (2021b, 2022a, 2022b, 2022c). However, as provided above,

[1] See Kapetanios and Papailias (2018), and the references therein, for a brief review of the big data literature in economics as well as a review of the appropriate methodologies which can handle large data.

[2] This mainly refers to the monthly VAT indicators first published in 22/12/2017. See, e.g., Kapetanios and Papailias (2021a).

the majority of such datasets have been available only for a relatively short amount of time (e.g., 2018 or later). This limited availability of real-time, and in some cases big data-based, indicators also holds in most developed economies. For example, for the US we also have indicators which are published with relatively short history; e.g., the Chicago Fed retail trade summary (weekly, January 2018), the US DOT traffic volume (weekly, December 2019), the Indeed job postings indicator (daily, February 2020), the Fed ETF purchase (daily, May 2020), the Federal Reserve Bank of Cleveland inflation nowcasting (daily, January 2022), among others. Similarly, for Germany we have series such as EC DG TAUXD imports and exports (weekly, June 2014), Eurocontrol air traffic (daily, January 2019) and Indeed job postings (daily, February 2020), among many others, which have short time history.

The motivation for this research is that, in such cases, the applied researcher is facing with a very difficult empirical dilemma. On one hand, we ideally want to construct a set of variables which are able to track the economic cycle in various frequencies and, are also able to provide real-time information of external (or internal) shocks to the economic system or potential changes in its structure. On the other hand, we require to have as much information as possible so that our estimation has smaller bias; no need to mention the "curse-of-dimensionality" where a large selection of variables (i.e. large variable dimension) would ideally need to have a long time series history to allow the required degrees of freedom.[3] Thus, should we sacrifice a number of indicators with short history in favour of those with more data availability and use a dataset with less variables? Or should we simply use the "bottom" part of our dataset, resulting in a full set of variables but with smaller history?

In the first case, our estimation is expected to exhibit smaller bias but our set of variables is probably not including those real-time indicators since they do not have long history. In the second case, we have the benefit of including those real-time indicators, and therefore the potential information they carry, but we have to be extremely cautious with our estimation techniques. This study attempts to provide a solution to the above dilemma by suggesting a novel approach based on multiple

[3]However, modern state-of-the-art machine learning approaches can overcome the problem of dimensionality; e.g., penalised regressions, random forests, support vector regressions, etc.

regressions with auxiliary variables.

We contribute to the applied literature with an approach which is simple and intuitive and, perhaps, it could be placed in the middle of the previously mentioned dilemma. In particular, we suggest to estimate a model with a "backcalculated" version of the real-time indicators to match the time dimension of the longer regressors. This can be done in three steps: (i) in the first step, we regress the variables with the short history on a number of auxiliary variables which are available for the entire period under study, (ii) then, we use these parameter estimates and "backcalculate"[4] those variables and obtain an estimated version which spans throughout the period under study, (iii) finally, we regress the original target variable on this nowfull set of variables and obtain the coefficient estimates.[5]

The term "backcalculation" is not coined here for the first time. Barbieri et al. (1999) is one of the first papers dealing with the term "backcalculation" or "retropolation". That report details the methodological steps followed by ISTAT in the production of backcalculated series. The main approach used is the Chow and Lin (1971) methodology used for time disaggregation and extrapolation of the base series. In the same line, Bracci et al. (1999) summarises the main guidelines followed by ISTAT in the retropolation of National Accounts data backwards to 1988. Following an approach similar to Barbieri et al. (1999), Di Fonzo (2003) extends the method of temporal disaggregation by Chow and Lin (1971) to backcasting high-frequency data using dynamic models. He uses a simple dynamic model to estimate missing past quarterly values for which only annual benchmarks are available by using the

[4]The term "backcalculation"in our context can also be referred to as "reverse extrapolation" or "reverse extension". It should not be confused with the process "backcasting", even though they share some similarities in principles, since "backcasting" is an attempt to use as much present information to obtain a better estimate of the historical values. In our case, we are not looking for a "good estimate" of the historical values of the short time series; instead, we are estimating this historical information only as an intermediate step. Our main target is to obtain better estimates in the final regression model.

[5]Some might be confused with the daily and weekly frequencies of the previously mentioned variables and consider the existence of a mixed-frequency issue. Mixed-frequency is not analysed at all in the paper and could be considered as an extension of the hereby suggested approach. In what follows, we prefer to keep our motivation simple and intuitive and assume that the dependent variable, its regressors and the auxiliary variables for the regressors are all observed in the same frequency. Empirically, this is something which can be easily done with bridge equations.

available, more recent quarterly series and one or more related series covering the quarters for which the back-estimated (or interpolated) values are desired. Caporin and Sartore (2006) is an early paper concerned with the theoretical and operational framework for estimating past values of relevant time series starting from a limited information set. Starting from a general approach to address relevant problems, their discussion points to the fact that linear models could be the preferred choice. They organise their approach in four steps: planning of the empirical analysis, data availability and collection, model selection (strategy) and production of the final estimates. This process is illustrated empirically with the backcalculation of the EU15 Industrial Production Index. We differentiate ourselves from the above literature in two ways: (i) first, all the previous studies, as well as those which belong to the backcasting literature, approach the "backcalculation" from either a univariate time series perspective or as aggregation of disaggregated variables problem, and (ii) none of the above papers provides estimates of the missing history of a time series based on regression of auxiliary variables.

This paper provides ample evidence in favour of the suggested approach through a large number of simulations. In particular, we include simulations with various values for the R^2 both in the auxiliary and main regression and we also use many sample sizes and many sizes for the data availability, or -better- the non-availability, of the time series with the short history. The comparison of the estimation performance in each case is done via means of Mean Squared Error (MSE) of the parameter vector.

Furthermore, we present the empirical results in the nowcasting of three target variables: (i) the consumer price index (CPI), (ii) the industrial production, and (iii) the unemployment rate. We use the weekly fuel volume and online job advertisements as our main predictors which are available only from 2018. We keep our empirical cases simple to allow full intuition and interpretation of the results. Our findings suggest that the suggested approach yields to improved parameter estimation which, in turn, result in better nowcasts.

The rest of this study is organised as follows. Section 2 describes the suggested approach and presents the two competing models, Section 3 discusses the Monte Carlo output, Section 4 presents the empirical evidence and, finally, Section 5 pro-

vides some concluding remarks and ideas for future work.

2 Backcalculation with Auxiliary Regressors

2.1 Single Regressor

Consider the classical linear regression model with y_t denoting the dependent variable observed during $t = 1, .., T$ and x_t denoting a single explanatory variable with a shorter history than y_t observed during $t = (T_a + 1), .., T$ with $T_a > 0$. Obviously, in the ideal case where $T_a = 0$ both time series have an equal number of observations (full history) for the entire $t = 1, .., T$ and the model to be estimated is:

$$\textbf{Model 0: } y_t = \beta x_t + \varepsilon_t,\ t = 1, .., T \tag{1}$$

with $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ and $\sigma_\varepsilon^2 < \infty$. In this case, the OLS estimator is:

$$\hat{\beta}^0 = \frac{\sum_{t=1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^{T}(x_t - \overline{x})^2}. \tag{2}$$

We can further simplify the above case and standardise our variables to have zero mean and unit variance. Then,

$$\hat{\beta}^0 = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}. \tag{3}$$

Continuing the discussion of Section 1, consider that x_t is an indicator where $T_a > 0$. Therefore, Model 0 and, hence, $\hat{\beta}^0$ are not applicable. For the shorter (overlapping) part of the time series history, $t = (T_a + 1), .., T$, both y_t and x_t are observed and, therefore, a reasonable choice would be to estimate:

$$\textbf{Model 1: } y_t = \beta x_t + \varepsilon_t,\ t = (T_a + 1), .., T. \tag{4}$$

In this case, the (lower variance) OLS estimator is:

$$\hat{\beta}^1 = \frac{\sum_{t=T_a+1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=T_a+1}^{T}(x_t - \overline{x})^2} \tag{5}$$

and, in the case of standardised variables:

$$\hat{\beta}^1 = \frac{\sum_{t=T_a+1}^{T} x_t y_t}{\sum_{t=T_a+1}^{T} x_t^2}. \tag{6}$$

In this case, it is easy to see that:

$$\begin{aligned}\hat{\beta}^1 &= \frac{\sum_{t=T_a+1}^{T} x_t(\beta x_t + \varepsilon_t)}{\sum_{t=T_a+1}^{T} x_t^2} \\ &= \beta + \frac{\sum_{t=T_a+1}^{T} x_t \varepsilon_t}{\sum_{t=T_a+1}^{T} x_t^2}\end{aligned}$$

An alternative approach would be to estimate a regression where the missing part of x_t, i.e. for $t = 1, .., T_a$, has been somehow estimated. One obvious choice would be to do a backcasting exercise. In this case we could reverse the time series x_t for $t = T, T-1, .., (T_a+1)$, fit a univariate time series model, e.g., an ARMA(p,q) and obtain the recursive forecasts, now backcasts, for $h = 1, 2, .., T_a$. Then, we can reverse the series back where the initial values at the top have been backcasted. This approach would not be very problematic for small backcasting horizons (i.e. T_a very small). However, as we move to longer horizons, the backcasting ability of these models deteriorates (i.e. the exact analogy we have with forecasting in long horizons where the forecast error increases). Therefore, a case where $T = 50$ and $T_a = 30$ would perhaps suffer from large bias.

Our suggestion is to use an auxiliary regression for x_t, $t = (T_a + 1), .., T$, as an intermediate step using a set of variables. For reasons of simplicity, we consider one auxiliary variable, z_t, $t = 1, .., T$. Therefore, this variable has history equal to that

of y_t. Then, we can first estimate the following model:

$$x_t = \gamma z_t + v_t,\ t = (T_a + 1), .., T \tag{7}$$

with v_t being independent from ε_t and $v_t \sim WN(0, \sigma_v^2)$ and $\sigma_v^2 < \infty$. In this case, the OLS estimator for γ is:

$$\widehat{\gamma} = \frac{\sum_{t=T_a+1}^{T}(z_t - \overline{z})(x_t - \overline{x})}{\sum_{t=T_a+1}^{T}(z_t - \overline{z})^2}$$

and, in the case of standardised variables we have:

$$\widehat{\gamma} = \frac{\sum_{t=T_a+1}^{T} z_t x_t}{\sum_{t=T_a+1}^{T} z_t^2}. \tag{8}$$

Since z_t goes back to $t = 1$, one can use the entire z_t and $\widehat{\gamma}$ and obtain the following "backcalculated" version of x_t:

$$\widehat{x}_t = \widehat{\gamma} z_t,\ t = 1, .., T_a. \tag{9}$$

Then, we can have the full version of x_t, denoted by $\widetilde{x}_t$ by concatenating $\widehat{x}_t$ which is backcalculated for $t = 1, .., T_a$ and x_t which is observed for $t = (T_a + 1), .., T$, i.e. $\widetilde{x}_t = (\widehat{x}_t, x_t)$ and estimate the following alternative model:

$$\textbf{Model 2: } y_t = \beta \widetilde{x}_t + \upsilon_t,\ t = 1, .., T. \tag{10}$$

where υ_t is independent from ε_t. In this case, the OLS estimator of β becomes:

$$\hat{\beta}^2 = \frac{\sum_{t=1}^{T}(\widetilde{x}_t - \overline{\widetilde{x}})(y_t - \overline{y})}{\sum_{t=1}^{T}(\widetilde{x}_t - \overline{\widetilde{x}})^2} \tag{11}$$

and, in the case of standardised variables we have:

$$\hat{\beta}^2 = \frac{\sum_{t=1}^{T} \widetilde{x}_t y_t}{\sum_{t=1}^{T} \widetilde{x}_t^2}. \tag{12}$$

We can then write:

$$\begin{aligned}\hat{\beta}^2 &= \frac{\sum_{t=1}^{T_a} \widehat{x}_t y_t}{\sum_{t=1}^{T_a} \widehat{x}_t^2} + \frac{\sum_{t=T_a+1}^{T} x_t y_t}{\sum_{t=T_a+1}^{T} x_t^2} \\ &= \frac{\sum_{t=1}^{T_a} \widehat{x}_t y_t}{\sum_{t=1}^{T_a} \widehat{x}_t^2} + \hat{\beta}^1\end{aligned}$$

The purpose of this paper is to investigate via simulations if the MSE of $\hat{\beta}^2$ is smaller than the MSE of $\hat{\beta}^1$ and the conditions under which this relationship holds. In both cases it is important to consider that the R^2 plays a key role as it indicates the percentage of the variability of the regressand explained by the variability of the regressor. Naturally, we have the R^2 in the main equation between y_t and x_t, denoted by R_y^2, and the R^2 in the auxiliary equation between x_t and z_t, denoted by R_x^2.

2.2 Multiple Regressors

We can generalise the approach described in the previous section by considering a large collection of x_t variables with a shorter history observed during $t = (T_a + 1), .., T$ with $T_a > 0$. As above, in the ideal case where $T_a = 0$ all time series have an equal number of observations (full history) for the entire $t = 1, .., T$ and the model to be estimated becomes:

$$\textbf{Model 0: } y_t = \beta_1 x_{1t} + ... + \beta_N x_{Nt} + \varepsilon_t,\ t = 1, .., T. \tag{13}$$

Since the above model is not applicable when the various regressors are not observed for the full time period, i.e. $T_a > 0$, we can consider the model which estimates the parameters in the common history part, i.e. $t = (T_a + 1), .., T$. The model now

becomes:

$$\textbf{Model 1: } y_t = \beta_1 x_{1t} + ... + \beta_N x_{Nt} + \varepsilon_t,\ t = (T_a + 1), .., T. \tag{14}$$

In the suggested approach, consider that there exist K auxiliary variables to be used in the intermediate step regression for each of the N regressors. Therefore, we can now write:

$$x_{nt} = \gamma_{n1} z_{1t} + ... + \gamma_{nK} z_{Kt} + \upsilon_t,\ t = (T_a + 1), .., T \tag{15}$$

for $n = 1, .., N$. Then, we can backcalculate each regressor as:

$$\widehat{x}_{nt} = \widehat{\gamma}_{n1} z_{1t} + ... + \widehat{\gamma}_{nK} z_{Kt},\ t = 1, .., T_a \tag{16}$$

where $\widehat{\gamma}_{nk}$ denotes the OLS estimator for $n = 1, .., N$ and $k = 1, .., K$. Finally, we can obtain the approximate to the full version of each x_{nt}, denoted by $\widetilde{x}_{nt}$, by concatenating $\widehat{x}_{nt}$ which is backcalculated for $t = 1, .., T_a$ and x_{nt} which is observed for $t = (T_a + 1), .., T$, i.e. $\widetilde{x}_{nt} = (\widehat{x}_{nt}, x_{nt})$ and estimate the following alternative model:

$$\textbf{Model 2: } y_t = \beta_1 \widetilde{x}_{1t} + ... + \beta_N \widetilde{x}_{Nt} + \upsilon_t,\ t = 1, .., T. \tag{17}$$

As in the previous case, it is important to highlight that effects from both R^2 coefficients in the main and auxiliary regressions, R_y^2 and R_x^2, should be thoroughly investigated.

3 Simulation Evidence

In this section of our research, we discuss the simulation evidence from our Monte Carlo experiments. In all cases, we consider the following sample sizes: $T = \{25,\ 50,\ 100,\ 200\}$. For each sample size, we consider the following "missing"/non-availability options for the time series with the shorter history: $T_a = \lfloor aT \rfloor$ where $a = \{0.05,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9\}$ and $\lfloor \cdot \rfloor$ denotes integer part. We also consider

the following options for the R^2 coefficients: $R_y^2 = \{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9\}$ and $R_x^2 = \{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9\}$. Finally, for the above we consider three main cases for N and K starting with the simplest one where $(N,\ K) = (1,\ 1)$ and further consider $(N, K) = \{(1,\ 3)\}$ and $(N, K) = \{(3,\ 3)\}$. Across all cases the coefficients are equal to unity, i.e. $\beta_1 = \beta_2 = \beta_3 = 1$ (where applicable) and $\gamma_1 = \gamma_2 = \gamma_3 = 1$ (where applicable).

Figures 1 to 12 report the relative MSE of Model 2 against the main benchmark which is Model 1 for all the above settings combinations. For a generic parameter θ and its corresponding estimator $\widehat{\theta}$, the MSE of $\widehat{\theta}$ is calculated as:

$$MSE(\widehat{\theta}) = Var(\widehat{\theta}) + Bias(\widehat{\theta}, \theta)^2$$

where $Var(\widehat{\theta}) = \frac{1}{R}\sum_{i=1}^{R}\left(\widehat{\theta}_i - \overline{\theta}\right)^2$ and $Bias(\widehat{\theta}, \theta) = \frac{1}{R}\sum_{i=1}^{R}\left(\widehat{\theta}_i - \theta\right)$ calculated across the number of Monte Carlo simulations R.

In our case, we compute:

$$MSE(\widehat{\beta}^1) = Var(\widehat{\beta}^1) + Bias(\widehat{\beta}^1, \beta)^2$$

and

$$MSE(\widehat{\beta}^2) = Var(\widehat{\beta}^2) + Bias(\widehat{\beta}^2, \beta)^2$$

for $R = 10,000$ Monte Carlo simulations. To facilitate the comparison of our results, we report the relative MSE defined as:

$$RelMSE = \frac{MSE(\widehat{\beta}^2)}{MSE(\widehat{\beta}^1)} \tag{18}$$

where $RelMSE < 1$ indicates that Model 2 improves against Model 1. In cases where we have $N > 1$, we report:

$$RelMSE = \frac{\sum_{n=1}^{N} MSE(\widehat{\beta}_n^2)}{\sum_{n=1}^{N} MSE(\widehat{\beta}_n^1)} \tag{19}$$

across the N parameters.

3.1 $N = 1,\ K = 1$

We start our simulations evidence with the simplest case considering the model described in Equation (1) with the sample sizes, T, non-availability periods, T_a, and R^2 coefficients as discussed above. To simplify the comparison of results, we consider $\beta = 1$ in the true model. Figures 1 to 4 display the relative MSE results for the choices of $T = \{25,\ 50,\ 100,\ 200\}$ respectively.

Starting with the smaller sample size of $T = 25$ in Figure 1, we see that when the R_y^2 is relatively low, i.e. up to about 40%, the suggested Model 2 provides MSE which improves against the main Model 1. This is true for all values of R_x^2 and it further improves (i.e. $RelMSE \to 0$) as R_x^2 increases. This is a reasonable, and expected, result as higher levels of R_x^2 indicate that the auxiliary model has better fit and, hence, the backcalculated values have smaller bias.

As we move to higher R_y^2, e.g., 50% or 60%, the fit of the main model depends more on the variability of the regressors. Therefore, a good fit becomes more important now. This is evident as we see that the relative MSE is below unity for $R_x^2 > 0.3$ or $R_x^2 > 0.5$. Of course, this is also true for reasonable choices of the non-availability period, i.e. $T_a = \lfloor aT \rfloor$ with $0.05 \leq a \leq 0.4$. Still, if we consider the experiment in actual numbers we have that, in a dataset of 25 observations where about the first 10 are missing, Model 2 improves against Model 1 for reasonable (and relatively mild) values of R_x^2.

As R_y^2 increases to extremely high levels of 70% or 80%, the fit of the main regression heavily depends on the good backcalculated values of the regressors. This, in turn, heavily depends on the fit of the auxiliary model and, hence, the appropriate regressor z_t. For $R_y^2 = 0.7$ (or 70%) we see that Model 2 improves against Model 1 only for $R_x^2 > 0.7$ for T_a with $a < 0.8$. Similarly, for $R_y^2 = 0.8$ Model 2 improves against Model 1 only for $R_x^2 > 0.8$ for T_a with $a < 0.5$.

In the extreme case where $R_y^2 = 0.9$ (or 90%), Model 1 has a better performance and the two models tend to have equal ability as $R_x^2 \to 1$ and low values of T_a.

Of course, in the previous discussion we have the case of a very small sample size of $T = 25$. Figure 2 presents the same output for the case of $T = 50$ where the overall qualitative conclusion does not change dramatically. We again see that for low values of R_y^2 Model 2 is better in all cases, however as R_y^2 increases Model 2 outperforms Model 1 for reasonable choices of R_x^2 and T_a. Figures 3 and 4 provide the corresponding results for $T = 100$ and $T = 200$.

Overall, this simulated experiment provides evidence in favour of the suggested approach. In models with low R_y^2 the researcher definitely needs to consider the suggested approach. In models with reasonable levels of R_y^2, e.g. 40% to 70%, the suggested approach should be considered if the auxiliary regression has a reasonable level of fit with R_x^2 in the range of 50% to 70%. Therefore, to apply this approach in practice, the researcher needs to seriously consider the fit of her main regression which, in the majority of macroeconomic nowcasting and forecasting with large datasets, it is not expected to be high.

3.2 $N = 1,\ K = 3$

We continue our experiment by including three regressors in Equation (15) with all other settings remaining identical to the previous case. The purpose of this exercise is to investigate if a larger number of auxiliary regressors affect the relative MSE of the two methods. Figures 5 to 8 present the results for the four sample size choices $T = \{25,\ 50,\ 100,\ 200\}$ respectively.

As expected, the results remain qualitatively similar to the previous case. However, we now see that as R_y^2 increases, we required larger values of R_x^2 in the auxiliary regression for Model 2 to outperform Model 1. Of course, in relatively mild values of R_y^2 and R_x^2, Model 2 outperforms Model 1 for various choices of T_a.

3.3 $N = 3,\ K = 3$

This is, perhaps, the most interesting case where we have three regressors in the auxiliary regression, i.e. $K = 3$ in Equation (15) and also three regressors in the main

equation, i.e. $N = 3$ in Equation (14) and Equation (17). Figures 9 to 12 present the results for the four sample size choices $T = \{25, 50, 100, 200\}$ respectively.

Starting with the case of the extremely small sample size, $T = 25$ in Figure 9, we see that now, for even large values of R_y^2, e.g. 0.4, 0.5, 0.6 or even 0.7, Model 2 should be preferred when (i) the non-available T_a part is large, i.e. $a > 0.7$, and/or R_x^2 takes reasonable values, i.e. $R_x^2 > 0.4$. In these cases, Model 2 dominates the benchmark. Even in the extreme case of $R_y^2 = 0.9$, where Model 2 did not outperform Model 1 in the previous cases, we now see that for $R_x^2 > 0.7$, Model 2 should be preferred.

The results which correspond to the larger sample sizes of $T = 50$, $T = 100$ and $T = 200$ further confirm the above arguments; see Figures 10 to 12. For relatively large values of R_y^2, say $R_y^2 = 0.6$, Model 2 outperforms Model 1 for $R_x^2 > 0.2$ indicating that, we overall have a better fit in the main model even if the fit in the auxiliary regression is low. In the case of $R_y^2 = 0.7$, Model 2 improves again Model 1 for $R_x^2 > 0.4$. In the more extreme case of $R_y^2 = 0.8$, Model 2 should be preferred when $R_x^2 > 0.6$.

Overall, the results in this case indicate that as we move to more complex DGPs, and perhaps closer to real applications where the true DGP remains unknown and a large number of variables should be considered, Model 2 has a lot of benefits to offer in the estimation of the coefficients in the main regression even for relatively low values of R_x^2. Model 2 further improves, and should be adopted, compared to Model 1 as the non-available part at the beginning of the time series is larger.

4 Empirical Evidence

4.1 Nowcasting Setup

Our empirical application attempts to compare the performance of Model 1 and Model 2 in terms of their nowcasting ability. In particular, we aim to answer the question: "does a longer -approximated- history of the predictors improve the estimation and, hence, reduce the bias in nowcasting?"

N1. For the first out-of-sample estimation, i, collect the corresponding data for y_t,

x_t and z_t (or z_{1t}, z_{2t}, etc.) for $t = 1, 2, ..., i$.

- y_t is expected to have some missing values due to publication lags.
- x_t is fully observed at time t.
- z_t might have publication lags or not (even if it does, it will not greatly affect our result as we mainly need it in the backcalculation of x_t).
- If x_t is of higher frequency than y_t (e.g., y_t could be monthly but x_t could be weekly), then the problem of mixed frequencies can be solved by Bridge equations.[6]

N2. Model 1 Nowcasts:

- Using the commonly observed history for y_t and x_t (both observed up to time i), we can estimate Model 1, $y_t = \beta x_t + \varepsilon_t$, or $y_t = \alpha + \beta x_t + \varepsilon_t$ if an intercept is allowed, and obtain $\widehat{\alpha}$ and $\widehat{\beta}$.
- Then, we can construct the nowcast estimate $\widehat{y}_i = \widehat{\beta} x_i$ or $\widehat{y}_i = \widehat{\alpha} + \widehat{\beta} x_i$ if an intercept is allowed.

N3. Model 2 Nowcasts:

- For the alternative model, we need to first estimate $x_t = \gamma_1 z_{1t} + \gamma_2 z_{2t} + ... + \gamma_K z_{Kt} + \upsilon_t$ using the commonly observed history up to time i and obtain the coefficient estimates.
- Then, we can construct the "backcalculated" version of x_t, $\widetilde{x}_t$, as $\widehat{x}_t = \widehat{\gamma}_1 z_{1t} + \widehat{\gamma}_2 z_{2t} + ... + \widehat{\gamma}_K z_{Kt}$ and obtain the full version of x_t by concatenating $\widehat{x}_t$ during the time that x_t is not observed and x_t for the periods which is actually observed.
- We can then estimate Model 2, $y_t = \beta \widetilde{x}_t + \upsilon_t$, or $y_t = \alpha + \beta \widetilde{x}_t + \upsilon_t$ if an intercept is allowed, and obtain $\widetilde{\alpha}$ and $\widetilde{\beta}$.

[6] An alternative approach could be to use the MIDAS or the U-MIDAS approaches, however it is not in our purpose to full examine the mixed frequency issue here.

- Finally, we can construct the nowcast estimate as $\widetilde{y}_i = \widetilde{\beta} x_i$ or $\widetilde{y}_i = \widetilde{\alpha} + \widetilde{\beta} x_i$ if an intercept is allowed.

N4. Go to Step N1 and repeat the above for all the out-of-sample estimation rounds.

By the end of the nowcasting exercise, we obtain the nowcast estimates for all the out-of-sample periods, T_{out}. To compare the predictive ability of each model we calculate the Mean Absolute Error (MAE) and the Root Mean Square Forecast Error (RMSFE) defined as:

$$MAE_m = \sqrt{\frac{1}{T_{out}} \sum_{j=1}^{T_{out}} |\widehat{e}_j|} \tag{20}$$

$$RMSFE_m = \sqrt{\frac{1}{T_{out}} \sum_{j=1}^{T_{out}} \widehat{e}_j^2} \tag{21}$$

for a generic model, Model m, where $\widehat{e}_j$ is the nowcast error for period j defined as $\left(y_j - \widehat{f}_j\right)$ where y_j is the actual value for the target and $\widehat{f}_j$ is the estimated nowcast. Then, we report the relative MAE and RMSFE defined as:

$$\text{Rel. } MAE = \frac{MAE_2}{MAE_1},$$
$$\text{Rel. } RMSFE = \frac{RMSFE_2}{RMSFE_1}.$$

In the above, a Rel. MAE or Rel. $RMSFE$ below unity indicates that the hereby suggested model, Model 2, has smaller error, and -thus- improved nowcasts, compared to the standard Model 1. We further provide the Diebold and Mariano (1995) p-value to highlight the statistical significance.

4.2 Cases

This section discusses the empirical output in terms of nowcasting for 12 cases across three target variables, y_t: (i) the consumer price index (CPI), (ii) the industrial

production, and (iii) the unemployment rate. In each of these cases, we consider variants of the weekly fuel volume and the online job advertisements as our predictor variables, x_t.[7] We further consider a number of auxiliary variables, z_t, which mainly include components of the unemployment rate, economic activity by sex, and the electricity and gas output. Table 1 provides more details for our variables. All data is publicly available by the ONS apart from the fuel volume which is published by the UK Department for Business, Energy & Industrial Strategy.

The monthly data for the target and auxiliary variables spans from 2009-12-31 to 2022-03-31 (148 monthly observations). The weekly data for the predictor variables spans from 2018-02-05 to 2022-03-28 (217 weekly observations). In our nowcasting exercise, we consider two evaluation periods: (i) *Eval. 1*: the first period spans from 2019-01-07 to 2020-01-27 (56 weekly out-of-sample periods/14 months) covering a relatively mild period of economic growth -which, however, includes the Brexit-related uncertainty-, and (ii) *Eval. 2*: the second period spans from 2019-01-07 to 2022-03-28 (169 weekly out-of-sample periods/about 3.5 years) which includes the COVID-19 pandemic.

4.2.1 Consumer Price Index

Tables 2 and 3 provide the results for the CPI in seven cases across the two evaluation periods. In each case, we report the relative MAE and RMSFE along with the p-value of the corresponding Diebold and Mariano (1995) test.

Case 1. In the first case, the target variable y_t is the consumer price index (gbpric00011). We start with only one predictor variable x_t which is the total fuel volume (gbcaes0150). For Model 2, we use one auxiliary variable z_t to backcast x_t which is the industrial production (gbprod0515). As we can see in Table 2, Model 2 has smaller nowcast error and significantly outperforms Model 1 in the relatively stable Eval. 1 period. The same holds in Eval. 2 period - even though the relative MAE slightly increases now.

Case 2. The second case builds on Case 1 and we attempt to use more aux-

[7] The mixed frequency issue is resolved by using Bridge equation.

iliary variables in an effort to provide better estimates which subsequently lead to better nowcasts. In this case we use two auxiliary variables z_t which are the industrial production (gbprod0515) and the production component of electricity, gas, etc. (gbprod0518). The results are similar to those of Case 1 without further improving the estimates; yet, they are still positive towards Model 2.

Case 3. The third case builds on Case 2 and we now try to predictor variables x_t which are the total fuel volume (gbcaes0150) and the fuel volume in London (gbcaes0144). The auxiliary variables z_t still remain the industrial production (gbprod0515) and the production component of electricity, gas, etc. (gbprod0518). Again, the results are in favour to Model 2.

It is important to notice at this point that the purpose of our research is not to provide the best nowcasts. Instead, we aim to offer robust empirical evidence in favour to the hereby suggested Model 2. Then, the applied researcher can use this as proof-of-concept and extensively search for the "best" auxiliary variables which minimise the nowcast error in a separate out-of-sample exercise or via the use of machine learning variable selection methods.

Case 4. The cases reported in Table 3 now use the online job advertisements as the predictor variables. In particular, we start by considering the target variable y_t to be the consumer price index (gbpric00011) as above. We start with only one predictor variable x_t which is the online job advertisements across all industries (gblama0103). We use four auxiliary variables z_t which are the unemployment rates of England (gblama1157), Northern Ireland (gblama1102), Scotland (gblama1048) and Wales (gblama0994). We can see that there is some evidence in favour to Model 2 in Eval. 1 but not in Eval. 2.

Case 5. We further attempt to improve the above case by considering two predictor variables x_t: the online job advertisements across all industries (gblama0103) and the online job advertisements in London (gblama0163) keeping the same auxiliary variables z_t. The results in Table 3 now show a great improvement. In Eval. 1, the relative MAE and RMSFE are 0.527 and 0.561 respectively while in the Eval. 2, which includes the COVID-19 pandemic, these statistics become 0.725 and 0.749.

Case 6. What would happen if we add more predictors in Case 5? Case 6

continues the same exercise but including three predictor variables x_t: the online job advertisements across all industries (gblama0103), the online job advertisements in London (gblama0163) and the online job advertisements in wholesale and retail (gblama0077) keeping the same auxiliary variables z_t. We now see a further significant decrease in the relative MAE and RMSFE indicating that the nowcast error of Model 2 is much smaller than that of Model 1. We now see in Table 3 a MAE and RMSFE of 0.350 and 0.402 for Eval. 1 which become 0.599 and 0.559 in Eval. 2 respectively.

Case 7. Finally, Case 7 builds on Case 6 and uses four predictor variables x_t: the online job advertisements across all industries (gblama0103), the online job advertisements in London (gblama0163), the online job advertisements in wholesale and retail (gblama0077) and the online job advertisements in manufacturing (gblama0074) while keeping the same auxiliary variables z_t. The results are slightly improved compared to Case 6.

4.2.2 Industrial Production

As in the previous tables, Table 4 provides the results for the industrial production in two cases across the two evaluation periods. In each case, we report the relative MAE and RMSFE along with the p-value of the corresponding Diebold and Mariano (1995) test.

Case 8. In this case we set the target variable y_t to be the industrial production (gbprod0515). We start with one predictor variable x_t which is the online job advertisements across all industries (gblama0103). For Model 2, we use four auxiliary variables z_t which are the unemployment rates of England (gblama1157), Northern Ireland (gblama1102), Scotland (gblama1048) and Wales (gblama0994). As we see, the Model 2 does a better nowcasting in both Eval. 1 and Eval. 2 periods; this outperformance is significant as indicated by the DM test.

Case 9. In this case keep the same target and auxiliary variables but we try four predictor variables which are the total fuel volume in England (gbcaes0143), Scotland (gbcaes0147) and London (gbcaes0144). The results in the stable Eval.

1 period do not indicate any statistically significant nowcasting gains in favour to Model 2, however the results in the Eval. 2 show a better performance of Model 2 with a relative MAE and RMSFE of 0.824 and 0.860 respectively.

4.2.3 Unemployment Rate

Table 5 provides the results for the unemployment rate in three cases across the two evaluation periods. In each case, we report the relative MAE and RMSFE along with the p-value of the corresponding Diebold and Mariano (1995) test.

Case 10. In this case we set the target variable y_t to be the unemployment rate (gblama00081). We start with two predictor variables x_t which are the online job advertisements across all industries (gblama0103) and the online job advertisements in London (gblama0163). For Model 2, we use two auxiliary variables z_t which are the economically active female percentage (gblama03961) and the economically active male percentage (gblama03951). In both evaluation periods, we see robust statistical evidence in favour of Model 2; the relative MAE and RMSFE are statistically significant smaller than unity.

Case 11. Case 11 extends the settings of Case 10 by using the same predictor variables but three auxiliary variables z_t which are the economically active female percentage (gblama03961), the economically active male percentage (gblama03951) and the cost & worked hours (gblama00361). By adding this extra auxiliary variable, we see that the nowcasts of Model 2 improved resulting in a relative MAE and RMSFE of 0.716 and 0.634 in the Eval. 1 period and 0.895 and 0.903 in the more volatile Eval. 2 period.

Case 12. Case 12 extends the settings of Case 11 by using three predictor variables x_t which are the online job advertisements across all industries (gblama0103), the online job advertisements in London (gblama0163) and the online job advertisements in the wholesale and retail (gblama0077). The results are similar to those of Case 11 indicating no significant value by the added predictor.

5 Conclusions

In this study we consider a model where the time series of the regressor is shorter than the time series of the regressand. This case of time misalignment is frequently met when the applied researcher includes real-time or big data-based indicators which are now widely available for a large variety of economies but only for 2-3 years of historical information.

In particular, we consider two models: one which is estimated only in the overlapping (short) time period, and an alternative one where the non-available part of the regressor has been first estimated (backcalculated) using a set of auxiliary variables observed for the entire time period. We use simulations and shed light on the cases and conditions under which this alternative model outperforms, in terms of MSE, the model estimated in the common time period. Our results across the combinations of variables in the main and auxiliary regression, various sample sizes, various percentages of missing data and various levels of R^2 indicate that in more complex models, which in any case are closer to reality, the suggested approach outperforms the corresponding benchmark even in cases with large R^2. Therefore, the simulations-based evidence offered in this study highlights that this approach should be considered by the applied researcher in her empirical investigations.

Furthermore, we also consider the empirical results in the nowcasting exercise of three target variables: (i) the consumer price index (CPI), (ii) the industrial production, and (iii) the unemployment rate. We use the weekly fuel volume and online job advertisements as our main predictors which are available only from 2018. Our findings indicate that the hereby suggested approach yields improved parameter estimation which, in turn, results in better nowcasts.

Future work that the authors are aiming to develop includes theoretical underpinnings and an extension of the empirical results where the "best" auxiliary variables are automatically selected from a large pool.

References

- Barbieri, G., Di Palma, F., Massari, S., Moauro, F., and Savio, G. (1999). Backward calculation of quarterly national accounts in Italy. OECD Meeting of National Accounts Experts.

- Bracci, L., Di Leo, F., and Pisani, S. (1999). Retropolating italian annual national accounts data according to esa95. OECD.

- Caporin, M. and Sartore, D. (2006). Methodological aspects of time series backcalculation. University Ca'Foscari of Venice, Dept. of Economics Research Paper Series, 56/06.

- Chow, G. and Lin, A.-L. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. The Review of Economics and Statistics, 372–375.

- Di Fonzo, T. (2003). Constrained retropolation of high-frequency data using related series: A simple dynamic model approach. Statistical Methods and Applications, 12(1):109–119.

- Kapetanios, G., Papailias, F. (2018). Big Data & Macroeconomic Nowcasting: Methodological Review. ESCoE Discussion Paper 2018-12.

- Kapetanios, G., Papailias, F. (2021a). Investigating the predictive ability of ONS big data-based indicators. Journal of Forecasting, 41(2), 252-258.

- Kapetanios, G., Papailias, F. (2021b). UK Economic Conditions during the Pandemic: Assessing the Economy using ONS Faster Indicators. ESCoE Discussion Paper 2021-10.

- Kapetanios, G., Papailias, F. (2022a). Real Time Indicators during the COVID-19 Pandemic Individual Predictors & Selection. ESCoE Technical Report TR-15.

- Kapetanios, G., Papailias, F. (2022b). An Evaluation Framework for Targeted Indicators Aggregates vs. Disaggregates. ESCoE Technical Report TR-17.
- Kapetanios, G., Papailias, F. (2022c). A Quality Assessment Framework for Maintaining & Publishing New Indicators. ESCoE Technical Report TR-18

Tables

	Description	Frequency	Mcode	Source
Targets (y_t)	Consumer Price Index, Total, Index	Monthly	gbpric00011	ONS
	Industrial Production, Total (More Decimals), Constant Prices, SA, Index	Monthly	gbprod0515	ONS
	Unemployment, Rate, SA	Monthly	gblama00081	ONS
Predictors (x_t)	Total Fuel, Volume, Total	Weekly	gbcaes0150	BEIS
	Total Fuel, Volume, London	Weekly	gbcaes0144	BEIS
	Total Fuel, Volume, England	Weekly	gbcaes0143	BEIS
	Total Fuel, Volume, Scotland	Weekly	gbcaes0147	BEIS
	Online Job Advertisements, All Industries, Index	Weekly	gblama0103	ONS
	Online Job Advertisements, Wholesale & Retail, Index	Weekly	gblama0077	ONS
	Online Job Advertisements, Manufacturing, Index	Weekly	gblama0074	ONS
	Online Job Advertisements, London, Index	Weekly	gblama0163	ONS
Aux. (z_t)	Electricity, Gas, Water, Output, Constant Prices, SA, Index	Monthly	gbprod0518	ONS
	England, Unemployment Rate, SA	Monthly	gblama1157	ONS
	Northern Ireland, Unemployment Rate, SA	Monthly	gblama1102	ONS
	Scotland, Unemployment Rate, SA	Monthly	gblama1048	ONS
	Wales, Unemployment Rate, SA	Monthly	gblama0994	ONS
	Economic Activity, Female	Monthly	gblama03961	ONS
	Economic Activity, Male	Monthly	gblama03951	ONS
	Productivity, Costs & Hours Worked, Total	Monthly	gblama00361	ONS

Table 1: Variable details. The monthly data spans from 2009-12-31 to 2022-03-31 (148 monthly observations). The weekly data spans from 2018-02-05 to 2022-03-28 (217 weekly observations). All data has been collected using Macrobond.

	Case 1		Case 2		Case 3	
	Eval. 1	**Eval. 2**	**Eval. 1**	**Eval. 2**	**Eval. 1**	**Eval. 2**
MAE	0.921	0.936	0.938	0.944	0.899	0.938
DM	0.007	0.027	0.046	0.043	0.004	0.001
RMSFE	0.912	0.891	0.915	0.896	0.916	0.892
DM	0.029	0.040	0.018	0.041	0.006	0.023

Table 2: Nowcasting CPI: Cases 1-3. The monthly data spans from 2009-12-31 to 2022-03-31 (148 monthly observations). The weekly data spans from 2018-02-05 to 2022-03-28 (217 weekly observations). Eval. 1 spans from 2019-01-07 to 2020-01-27 (56 weekly out-of-sample periods $\approx$ 14 months). Eval. 2 spans from 2019-01-07 to 2022-03-28 (169 weekly out-of-sample periods $\approx$ 3.5 years). Reported are the relative MAE and RMSFE of Model 2/Model 1 and the p-values for the Diebold and Mariano (1995) test.

	Case 4		Case 5	
	Eval. 1	**Eval. 2**	**Eval. 1**	**Eval. 2**
MAE	0.824	0.979	0.527	0.752
DM	0.002	0.299	0.001	0.000
RMSFE	0.925	0.995	0.561	0.749
DM	0.219	0.782	0.006	0.000
	Case 6		**Case 7**	
	Eval. 1	**Eval. 2**	**Eval. 1**	**Eval. 2**
MAE	0.350	0.599	0.357	0.569
DM	0.000	0.000	0.000	0.000
RMSFE	0.402	0.559	0.314	0.450
DM	0.000	0.000	0.001	0.000

Table 3: Nowcasting CPI: Cases 4-7. The monthly data spans from 2009-12-31 to 2022-03-31 (148 monthly observations). The weekly data spans from 2018-02-05 to 2022-03-28 (217 weekly observations). Eval. 1 spans from 2019-01-07 to 2020-01-27 (56 weekly out-of-sample periods $\approx$ 14 months). Eval. 2 spans from 2019-01-07 to 2022-03-28 (169 weekly out-of-sample periods $\approx$ 3.5 years). Reported are the relative MAE and RMSFE of Model 2/Model 1 and the p-values for the Diebold and Mariano (1995) test.

	Case 8		Case 9	
	Eval. 1	**Eval. 2**	**Eval. 1**	**Eval. 2**
MAE	0.890	0.884	1.005	0.824
DM	0.165	0.000	0.960	0.000
RMSFE	0.844	0.945	0.973	0.860
DM	0.022	0.011	0.812	0.001

Table 4: Nowcasting Industrial Production: Cases 8-9. The monthly data spans from 2009-12-31 to 2022-03-31 (148 monthly observations). The weekly data spans from 2018-02-05 to 2022-03-28 (217 weekly observations). Eval. 1 spans from 2019-01-07 to 2020-01-27 (56 weekly out-of-sample periods $\approx$ 14 months). Eval. 2 spans from 2019-01-07 to 2022-03-28 (169 weekly out-of-sample periods $\approx$ 3.5 years). Reported are the relative MAE and RMSFE of Model 2/Model 1 and the p-values for the Diebold and Mariano (1995) test.

	Case 10		Case 11		Case 12	
	Eval. 1	**Eval. 2**	**Eval. 1**	**Eval. 2**	**Eval. 1**	**Eval. 2**
MAE	0.884	0.931	0.716	0.895	0.669	0.896
DM	0.042	0.003	0.017	0.003	0.000	0.000
RMSFE	0.808	0.926	0.634	0.903	0.669	0.926
DM	0.025	0.001	0.016	0.001	0.000	0.000

Table 5: Nowcasting Unemployment Rate: Cases 10-12. The monthly data spans from 2009-12-31 to 2022-03-31 (148 monthly observations). The weekly data spans from 2018-02-05 to 2022-03-28 (217 weekly observations). Eval. 1 spans from 2019-01-07 to 2020-01-27 (56 weekly out-of-sample periods $\approx$ 14 months). Eval. 2 spans from 2019-01-07 to 2022-03-28 (169 weekly out-of-sample periods $\approx$ 3.5 years). Reported are the relative MAE and RMSFE of Model 2/Model 1 and the p-values for the Diebold and Mariano (1995) test.

Figures

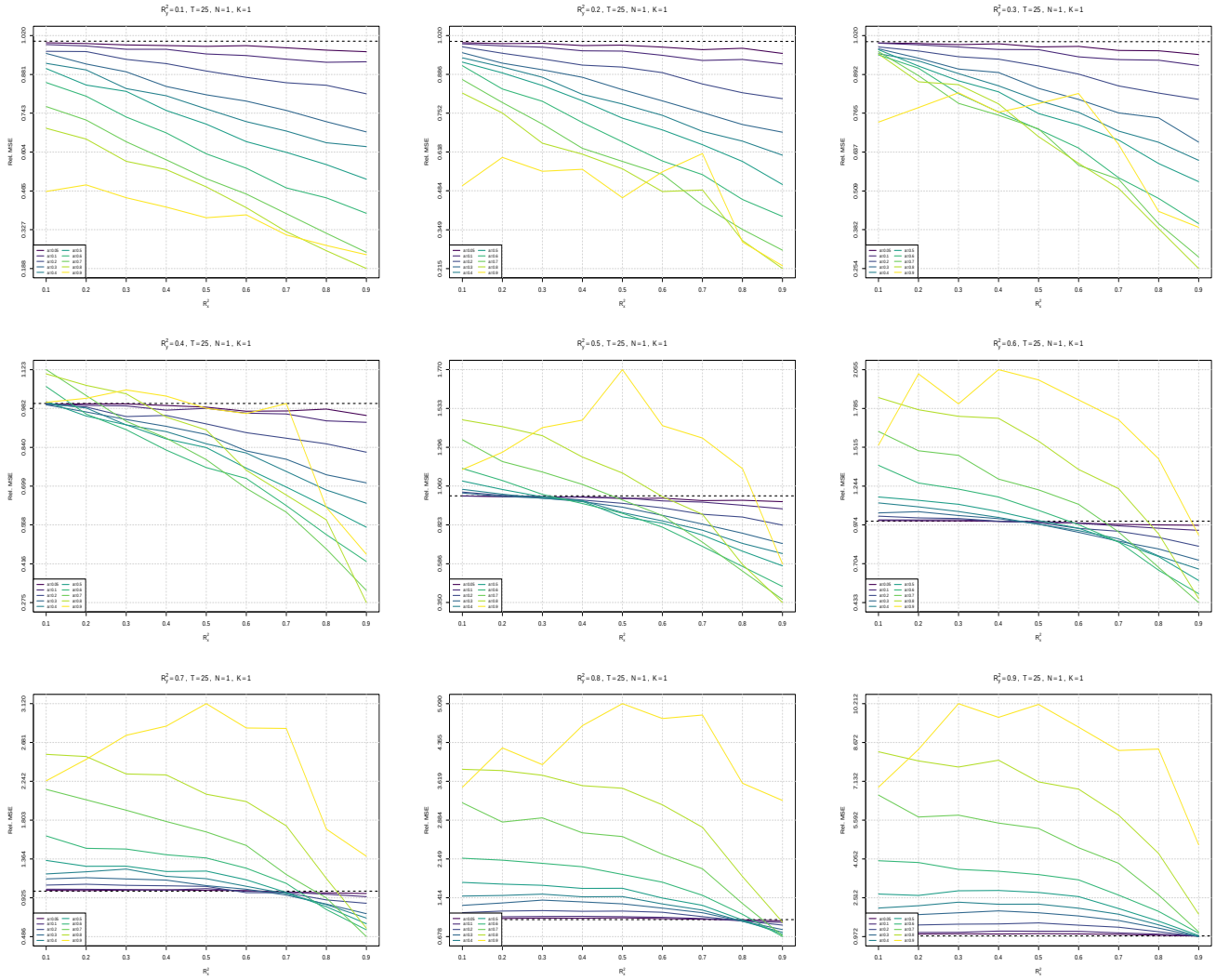

Figure 1: $N = 1$, $K = 1$, $T = 25$ with various T_a, R_y^2 and R_x^2.

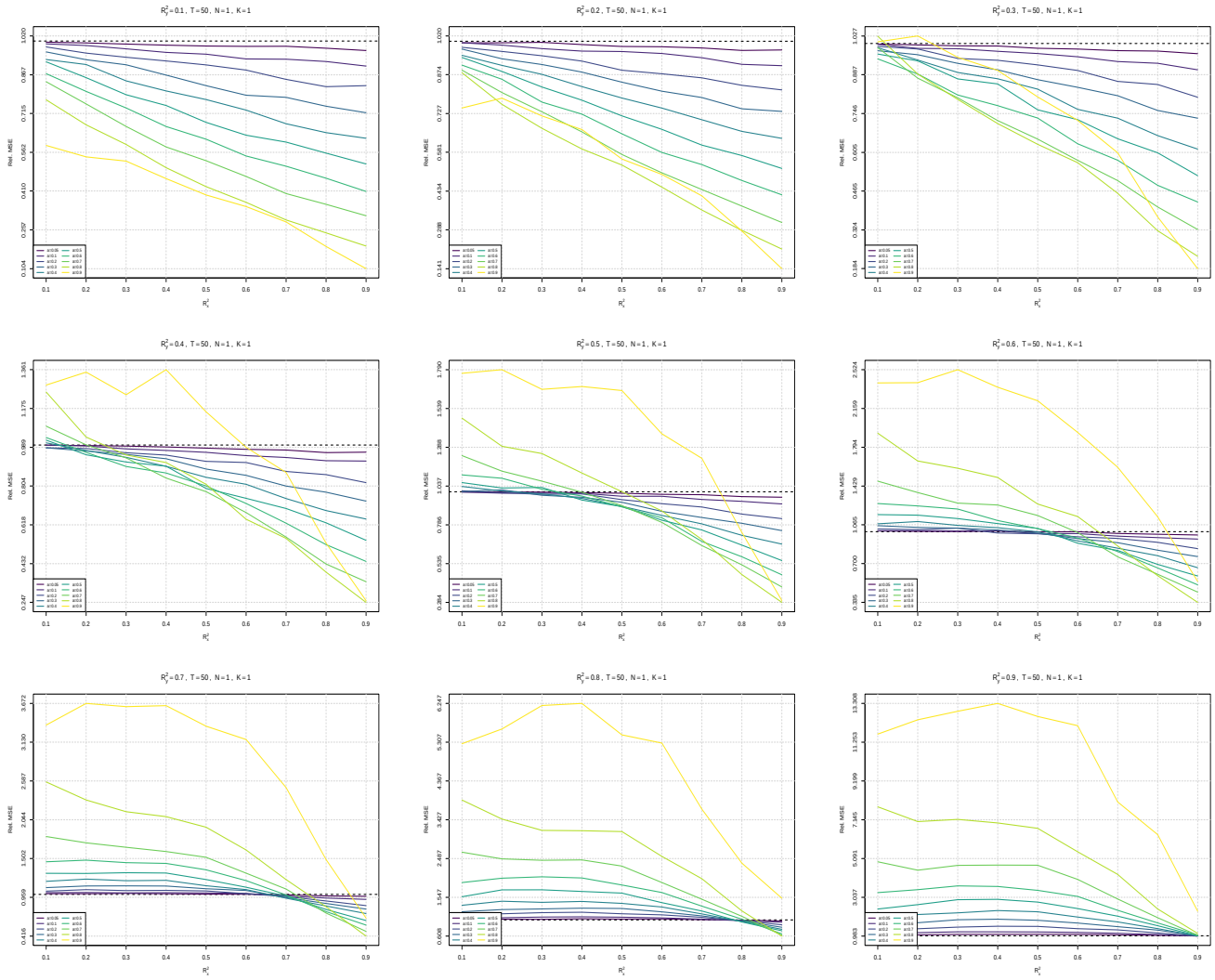

Figure 2: $N = 1$, $K = 1$, $T = 50$ with various T_a, R_y^2 and R_x^2.

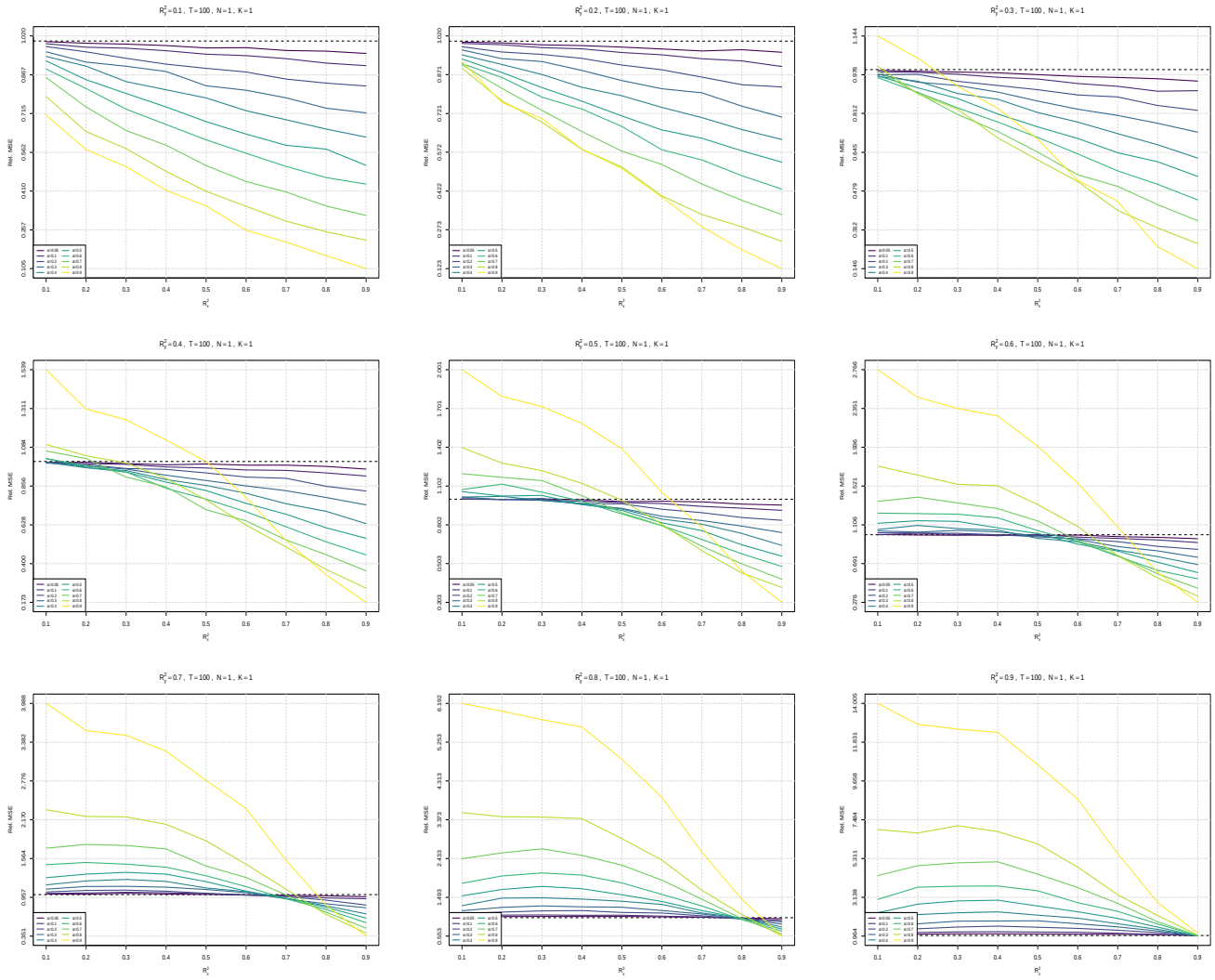

Figure 3: $N = 1$, $K = 1$, $T = 100$ with various T_a, R_y^2 and R_x^2.

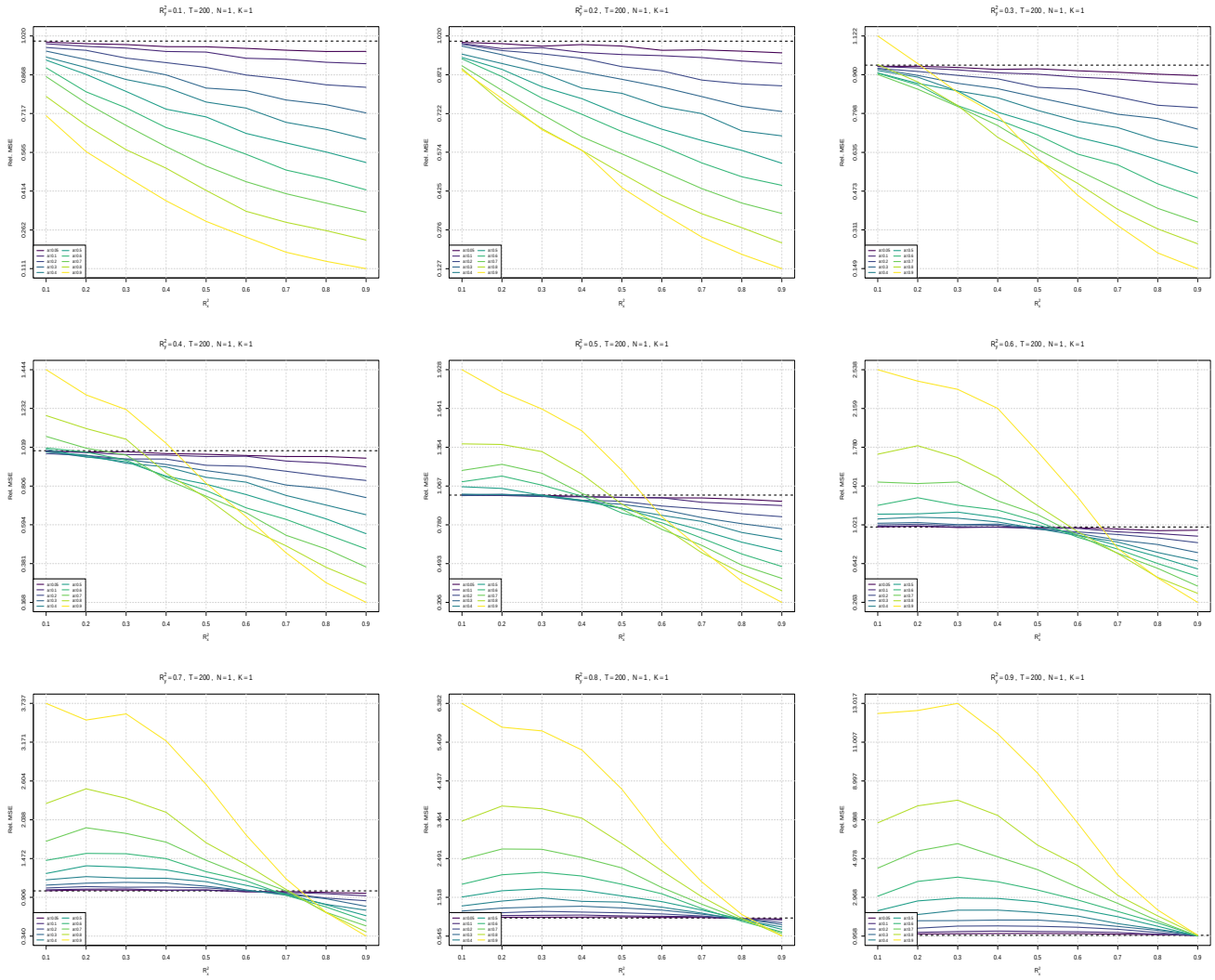

Figure 4: $N = 1$, $K = 1$, $T = 200$ with various T_a, R_y^2 and R_x^2.

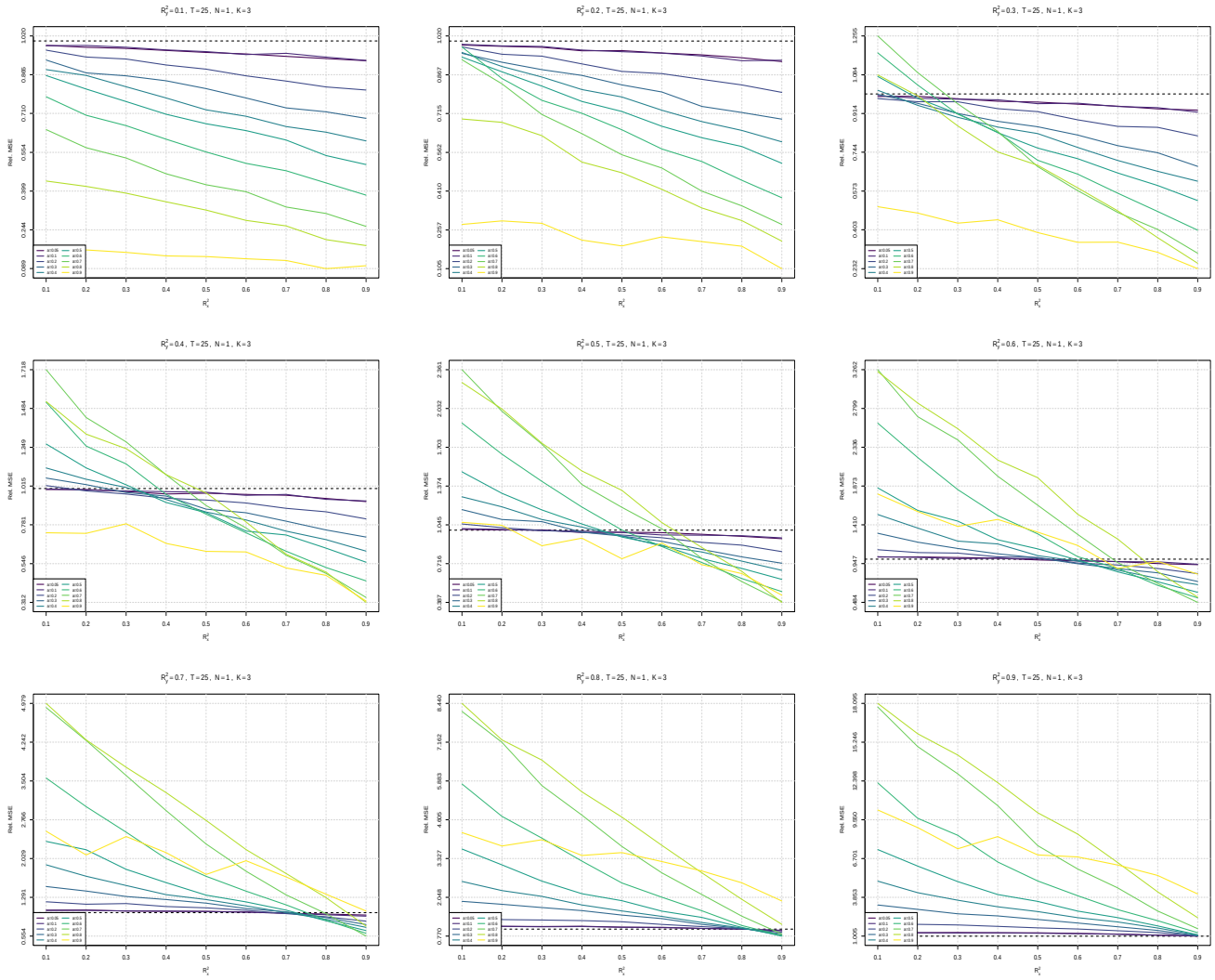

Figure 5: $N = 1$, $K = 3$, $T = 25$ with various T_a, R_y^2 and R_x^2.

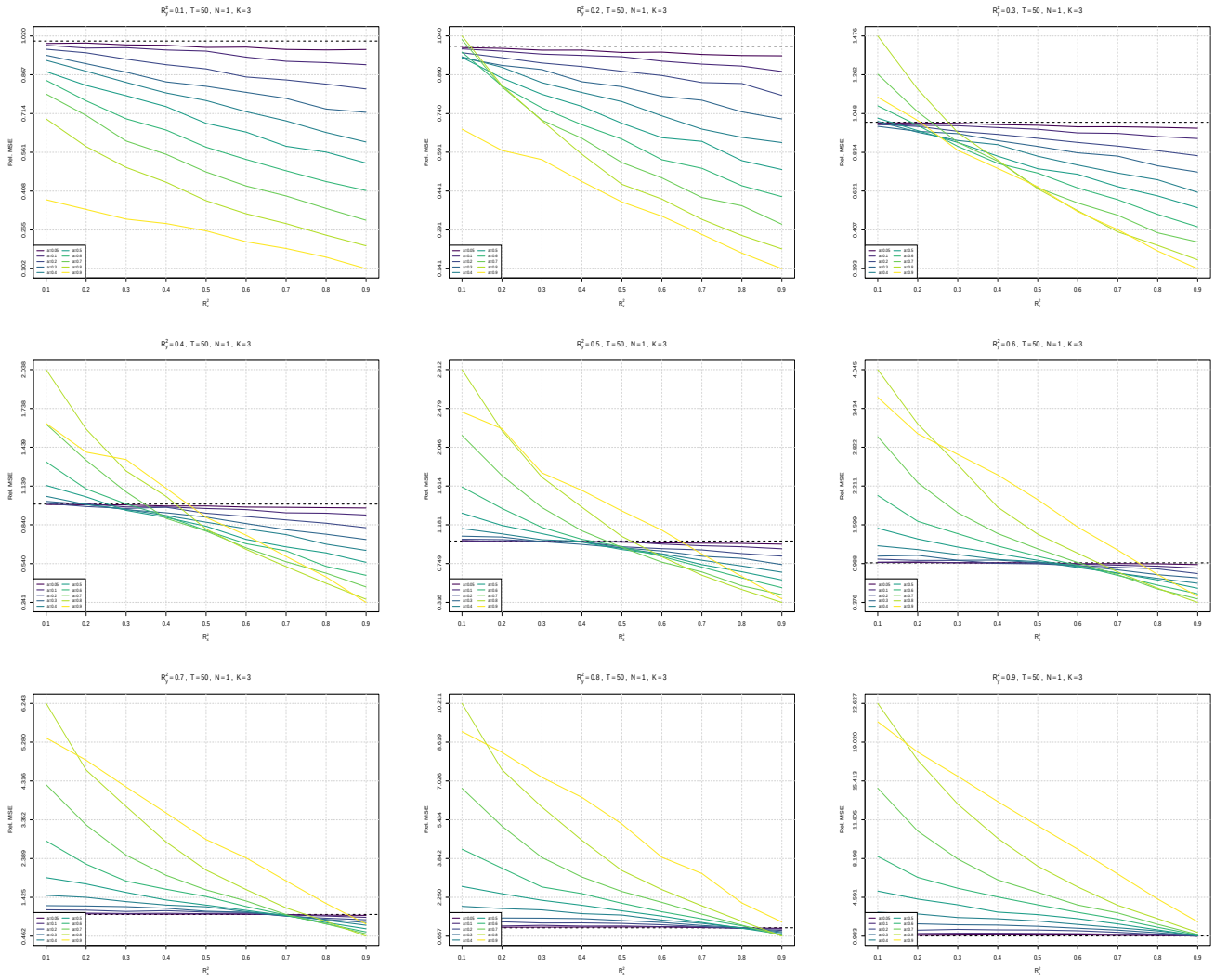

Figure 6: $N = 1$, $K = 3$, $T = 50$ with various T_a, R_y^2 and R_x^2.

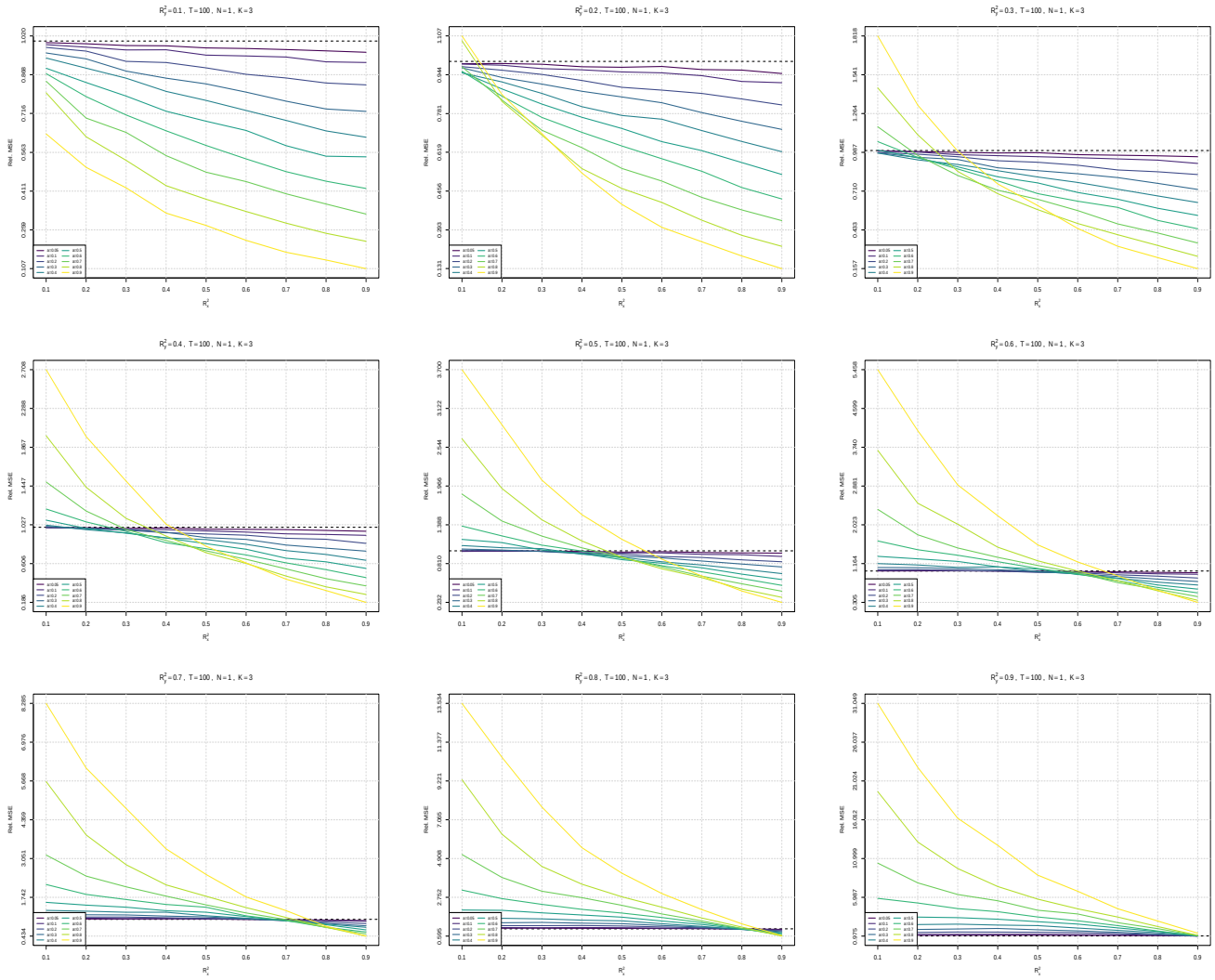

Figure 7: $N = 1$, $K = 3$, $T = 100$ with various T_a, R_y^2 and R_x^2.

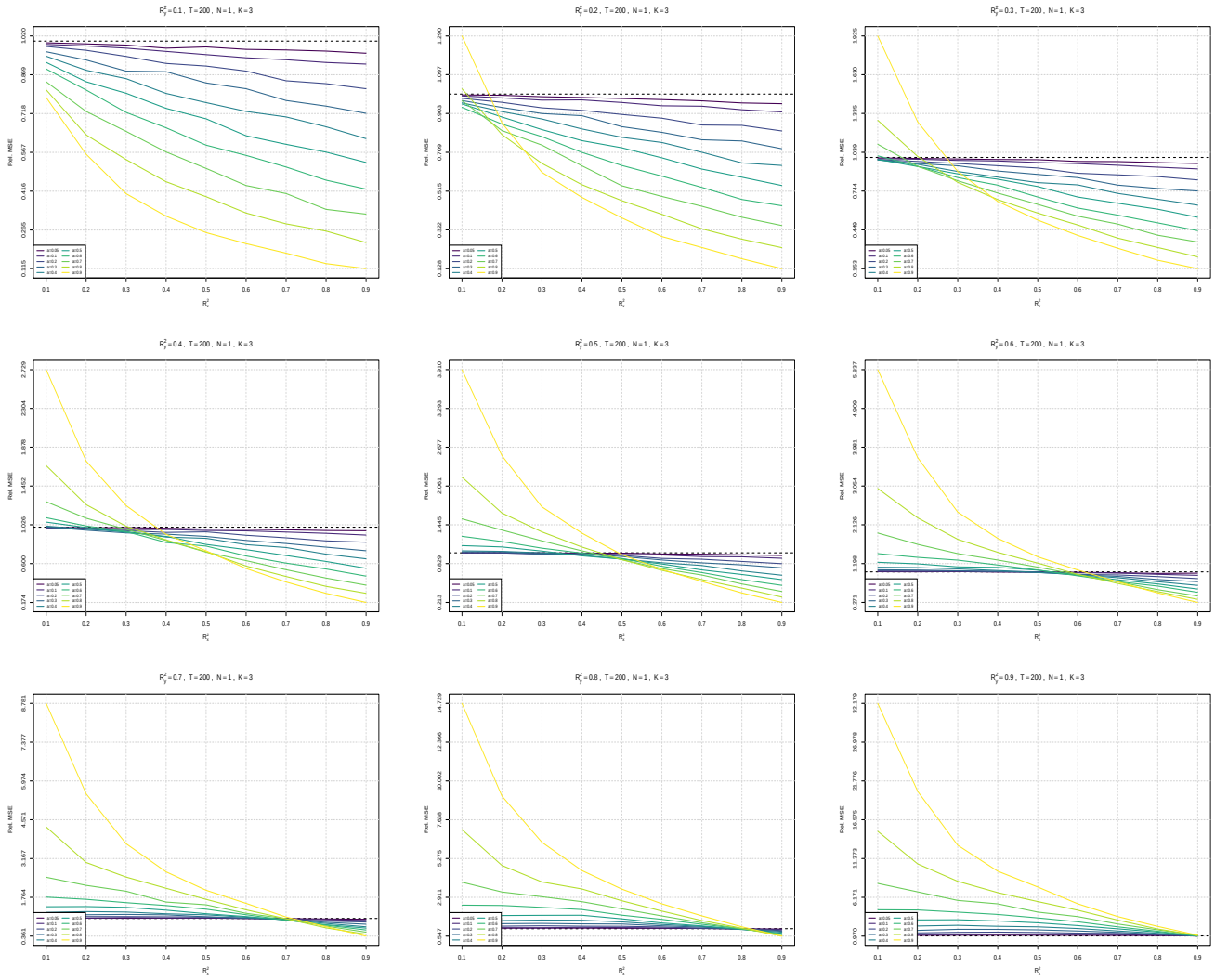

Figure 8: $N = 1$, $K = 3$, $T = 200$ with various T_a, R_y^2 and R_x^2.

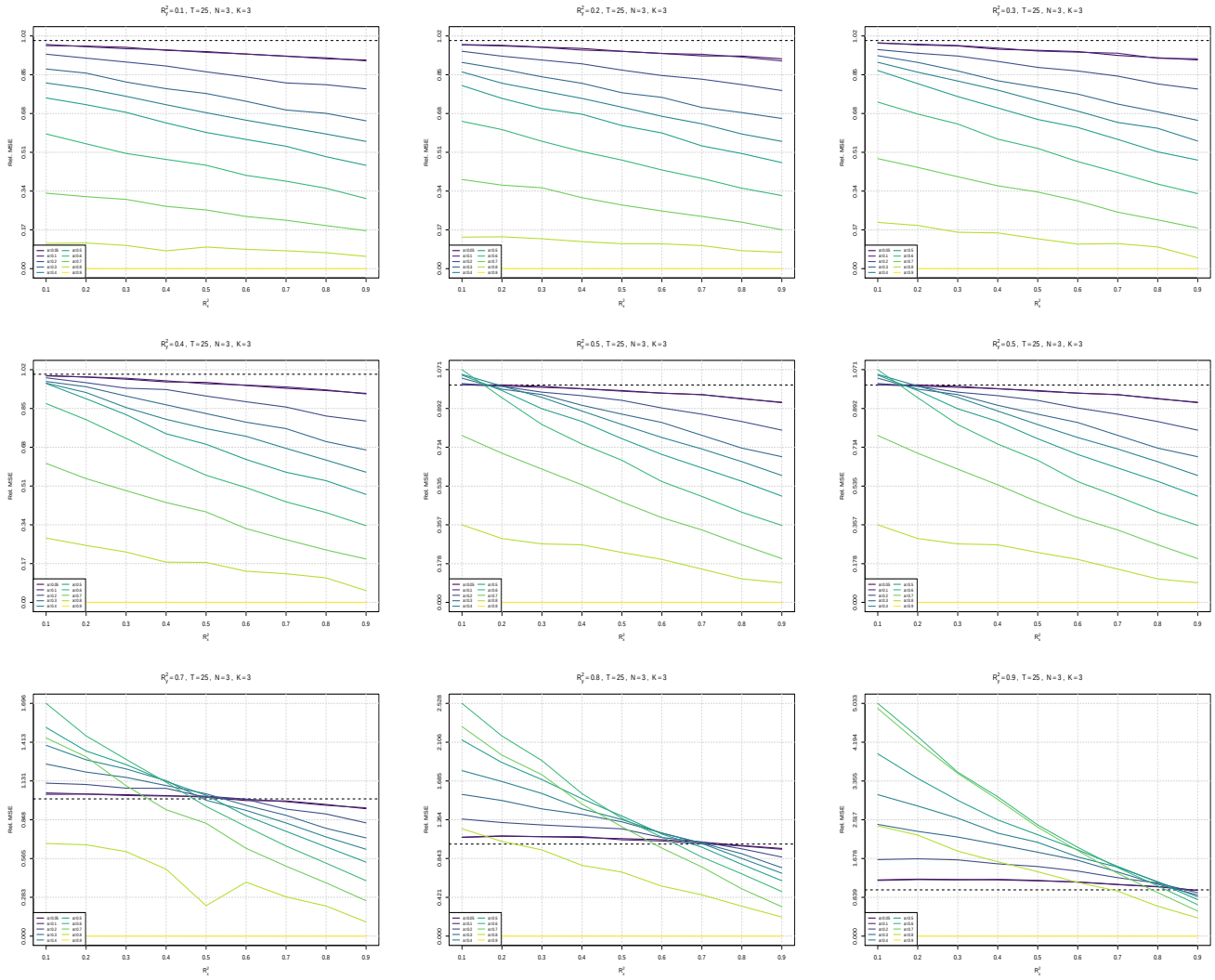

Figure 9: $N = 3$, $K = 3$, $T = 25$ with various T_a, R_y^2 and R_x^2.

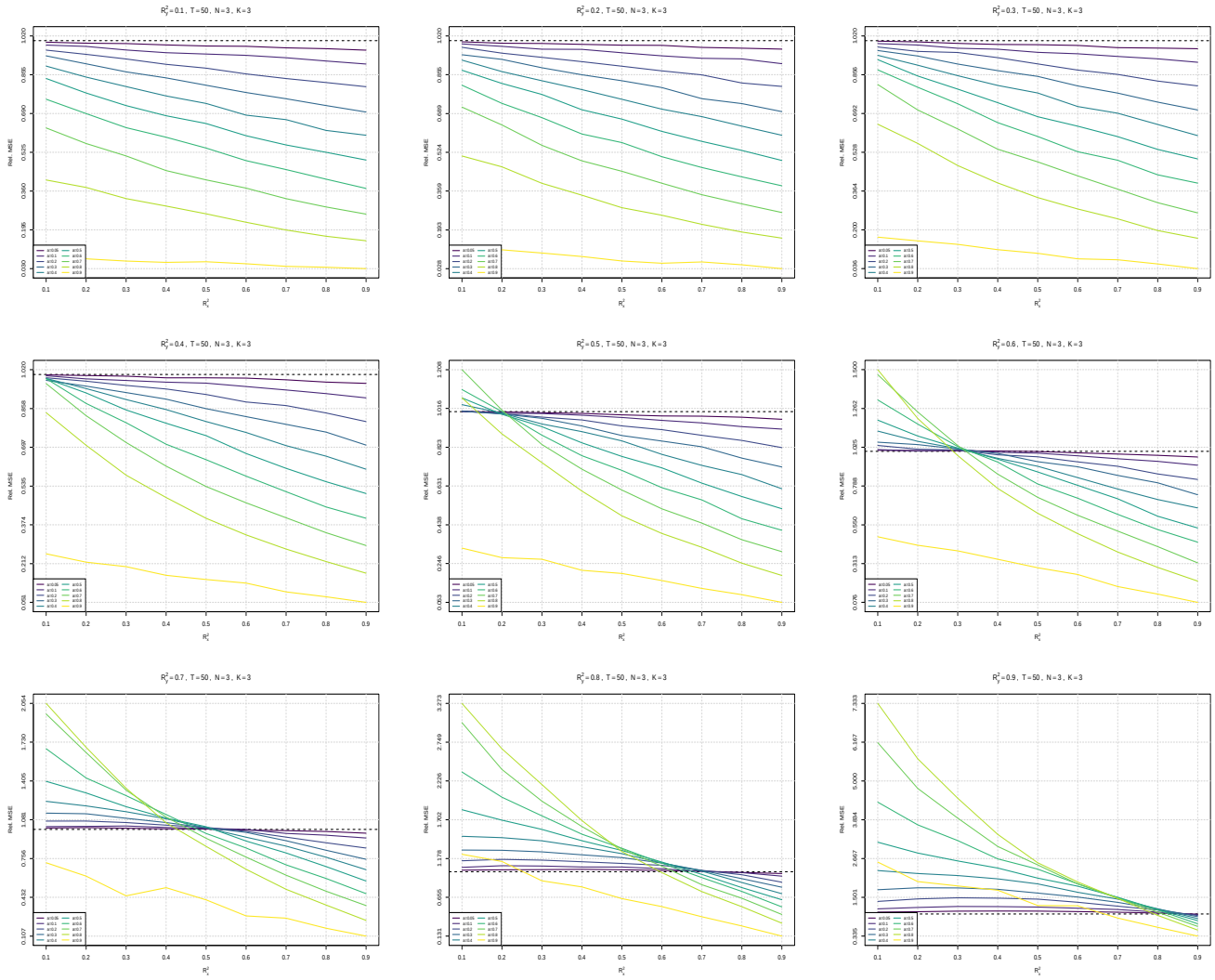

Figure 10: $N = 3$, $K = 3$, $T = 50$ with various T_a, R_y^2 and R_x^2.

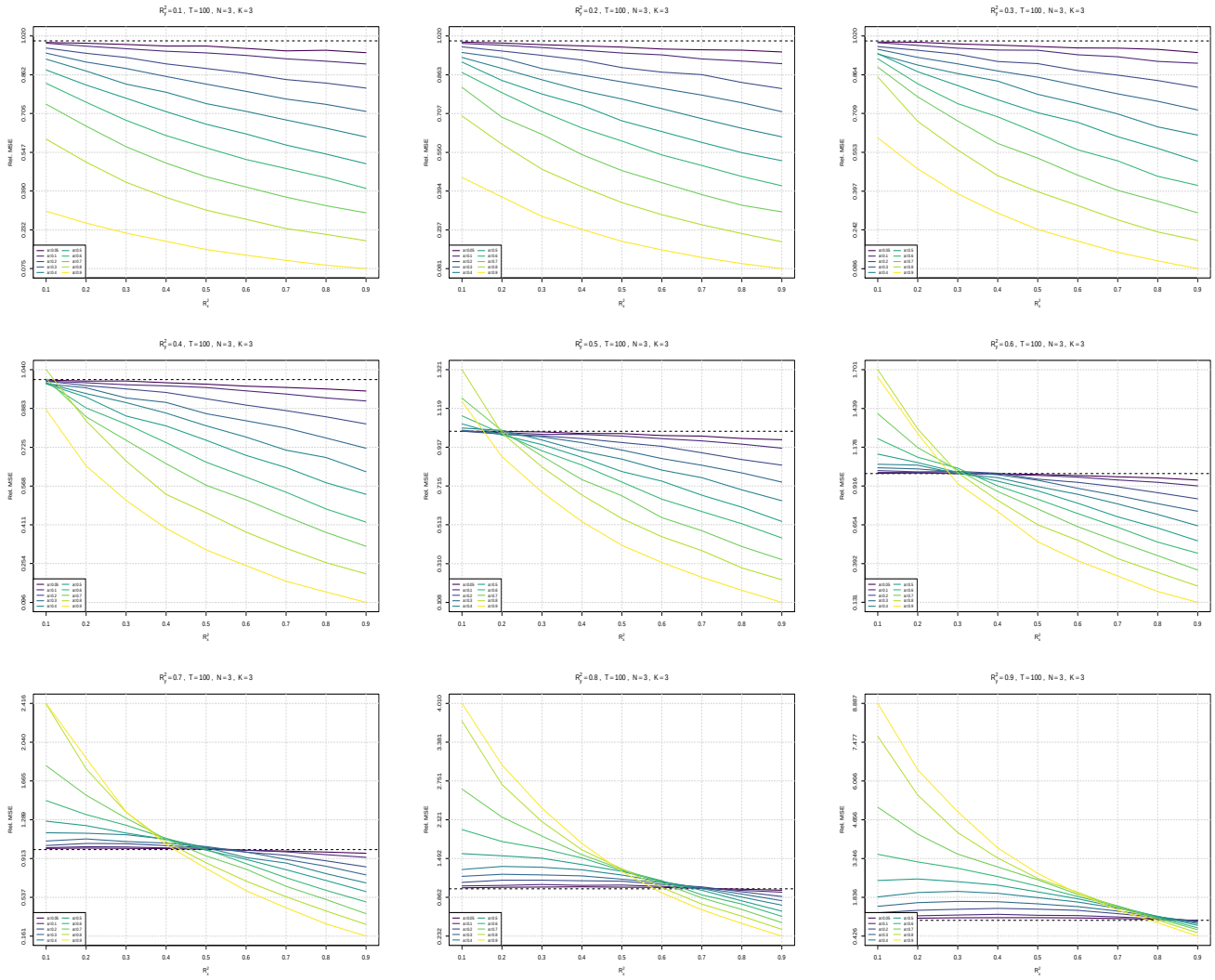

Figure 11: $N = 3$, $K = 3$, $T = 100$ with various T_a, R_y^2 and R_x^2.

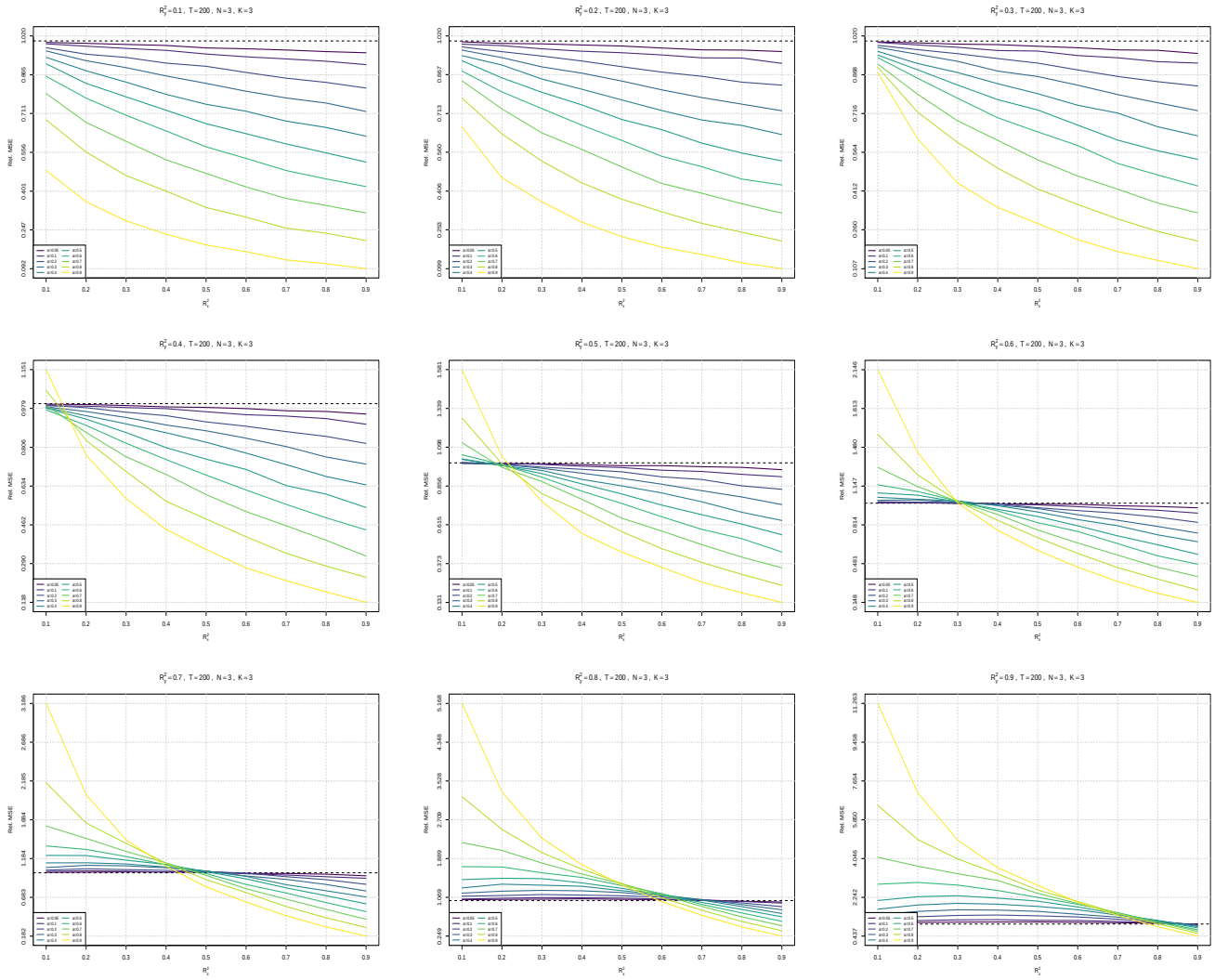

Figure 12: $N = 3$, $K = 3$, $T = 200$ with various T_a, R_y^2 and R_x^2.