

M-quantile regression analysis of temporal gene expression data

Veronica Vinciotti* & Keming Yu

Department of Mathematics, Brunel University,
Uxbridge UB8 3PH, UK

veronica.vinciotti@brunel.ac.uk & keming.yu@brunel.ac.uk

February 26, 2009

Abstract

In this paper, we explore the use of M-regression and M-quantile coefficients to detect statistical differences between temporal curves that belong to different experimental conditions. In particular, we consider the application of temporal gene expression data. Here, the aim is to detect genes whose temporal expression is significantly different across a number of biological conditions. We present a new method to approach this problem. Firstly, the temporal profiles of the genes are modelled by a parametric M-quantile regression model. This model is particularly appealing to small-sample gene expression data, as it is very robust against outliers and it does not make any assumption on the error distribution. Secondly, we further increase the robustness of the method by summarising the M-quantile regression models for a large range of quantile values into an M-quantile coefficient. Finally, we employ a Hotelling T^2 -test to detect significant differences of the temporal M-quantile profiles across conditions. Simulated data shows the increased robustness of M-quantile regression methods over standard regression methods. We conclude by using the method to detect differentially expressed genes from time-course microarray data on muscular dystrophy.

¹To whom correspondence should be addressed.

1 Introduction

Time-course gene expression data are often measured to study dynamic biological systems and gene regulatory networks. The data are produced by so called microarray experiments and they provide expression measurements for thousands of genes in a biological system under different time points and/or biological conditions (e.g. different diseases). To account for time dependency of the gene expression measurements over time and the noisy nature of microarray data, standard regression models such as mixed-effects models are normally used in analyzing these data (Luan & Li, 2003; Ng et al., 2006; Archer & Guennel, 2006; Ma et al., 2006; Ma & Zhong, 2008). However, these models have some drawbacks: they depend on strong distributional assumptions, they require a formal specification of the random part of the model and they ignore possible outliers. In this paper, we look at an alternative to these classical models when the aim is the identification of biologically interesting genes from temporal microarray data.

In general, the identification of biologically interesting genes in a temporal expression profiling dataset is challenging and complicated by high levels of experimental noise. A variety of methods have been suggested in the literature for the detection of differentially expressed genes. Only few of these deal with the most general situation where both temporal and biological conditions are present in the data (Park et al., 2003; Storey et al., 2005; Vinciotti et al., 2006; Storey et al., 2007; Yuan & Kendzierski, 2006). Recently, some methods have appeared which make use of quantiles and quantile regression models to detect differentially expressed genes (Wang & He, 2007, 2008; Yu et al., 2007). In (Yu et al., 2007), we discuss the advantages of using quantile regression over standard regression, especially when modelling gene expression data. In light of the positive results obtained in (Yu et al., 2007), in this paper we explore how M-quantile regression can be successfully used to model gene expression data.

M-quantile regression (Breckling & Chambers, 1988; Chambers & Tzavidis, 2006) combines the ideas of characterisation of the relationship between a response variable and explanatory variables when the behaviour of “non-average” individuals is of interest. The method of M-quantile regression is based on a “quantile-like” generalization of regression and influence function for M-estimation and as such provides a robust alternative to standard regression models. Indeed, this was proven on a small number of applications (Kokic & Chambers, 1997; Chambers & Tzavidis, 2006). In this paper, we extend both the model and the range of applications of this methodology.

The aim of this study is to test for differences in gene expression data across a number of conditions and time points. Similarly to Chambers & Tzavidis (2006), we use M-quantile coefficients to capture differences across the conditions. We first

derive a time-dependent conditional M-quantile coefficient based on a parametric polynomial model which takes also array effects into account. This results in the estimation of M-quantile coefficients profiles over time for each biological condition. A Hotelling T^2 -test is finally used to detect the genes which show significant difference in the temporal profiles across the conditions. Section 2 motivates the model, by comparing it with its standard alternatives, and derives the conditional M-quantile coefficients. A simulated example is provided in Section 3, while Section 4 discusses an interesting application to the analysis of muscular dystrophy gene expression data. Finally, in Section 5 we discuss the results and draw some conclusions.

2 The method

2.1 Why M-quantile?

It is well known how linear least-squares estimates can behave badly when the error distribution is not normal, particularly when the errors are heavy-tailed. To address this issue, two different approaches have been developed in the literature. One approach, termed robust regression, is to employ a fitting criterion that is not as vulnerable as least squares to unusual data. The most common general method of robust regression is M-estimation, introduced by Huber (1964). The second approach, which is also robust to large outliers, is the one of quantile regression (Koenker, 2005). In particular, quantile regression is used when the conditional variability across the population of interest can be characterised by the different quantiles of the population units.

The robustness of robust regression and quantile regression has been measured theoretically in terms of their breakdown point. Intuitively, the breakdown point of an estimator is the proportion of incorrect observations (i.e. arbitrarily large observations) an estimator can handle before giving an arbitrarily large result. The higher the breakdown point of an estimator, the more robust it is. According to this measure, it was found that robust regression, also called M-regression, has a high breakdown point (Mendes & Tyler, 1996), whereas quantile regression such as LAD (Least Absolute Derivation) regression, is not at all robust to observations with unusual predictor values; that is, it has a low breakdown point (Giloni et al., 2006). In view of these results, M-quantile regression provides a good alternative to both models. It integrates both M-regression and quantile regression, by providing a “quantile-like” generalization of regression based on influence functions. That is, it is as robust as M-regression to outliers and contaminated data, and it is used when the conditional variability across the population of interest can be characterised by

the M-quantile coefficients of the population units.

In this paper, we show how M-quantile regression can be used to detect differentially expressed genes via the use of M-quantile coefficients. These are described in the next section. The approach is particularly appealing for this application: first of all, it does not require strong distributional assumptions and it is more robust against outliers, which are expected in gene expression data due to the small sample size and the noisy nature of the data. Indeed, M-quantile models automatically provide robust inference against outliers, which is seldomly considered under the existing models for gene expression data. Secondly, in contrast to mixed-effects models for gene expression analysis, the application of M-quantile regression avoids the problems associated with the specification of random effects, allowing inter-condition differences to be characterised by condition-specific M-quantile coefficients. This is discussed also by Chambers & Tzavidis (2006) in their application to small-area estimation. Finally, the M-quantile coefficients summarise the information over all quantiles of the distribution, so a test based on these is expected to perform better than, for example, a median quantile regression, when measuring the conditional variability across the experimental conditions.

2.2 M-quantile coefficients

In this section, we derive the M-quantile coefficients and explore how these can be used to characterise the level of gene expression over time under different biological conditions. Loosely speaking, the M-quantile coefficient of a point (x, y) is the quantile value associated to the M-quantile regression line that passes through this point. In (Chambers & Tzavidis, 2006), an application of the M-quantile coefficients is described to small area estimation. There the main idea is that different areas correspond to different average values of the M-quantile coefficients. In this paper, we extend this idea to gene expression data, by taking into account both the temporal dimension of the data as well as the presence of different biological conditions. In general, we expect that the variability between biological conditions is a significant part of the overall variability of the gene population. That is, we expect genes under a particular condition to have similar M-quantile coefficients over time (although these coefficients may not be equal for all observed time points they are expected to be identical statistically).

Most of the papers on gene expression data either deal with the temporal modelling of gene expression data or with statistical tests for different biological conditions, but rarely combine these two. When both time and conditions are present in the data, instead of a single value of the M-quantile coefficient for each condition (as in (Chambers & Tzavidis, 2006)), the M-quantile coefficient now becomes a function of time. Our aim is then to estimate the M-quantile coefficient for differ-

ent time points and biological conditions and finally to develop a statistical test to detect differences in the temporal M-quantile coefficient profiles across biological conditions.

Let y_{tc} denote the gene expression at time t under the biological condition c and for a particular gene. Let μ_{ct} denote the true gene expression at time t and under condition c . The vector $\mu_c = (\mu_{c2}, \dots, \mu_{cT})$ for all time points t is the true temporal profile for a specific gene under the biological condition c . Similarly to (Vinciotti et al., 2006), we consider the following standard linear model

$$y_{ct} = \mu_{ct} + \eta_A + \epsilon_{ct}, \quad (2.1)$$

where η_A is the array effect, to account for the fact that gene expressions from the same array are normally more similar to each other than gene expressions from different arrays simply due to experimental errors, and ϵ_{ct} is the error term with mean zero and constant variance. Then we model the gene expression at time t for condition c , where $c = \{1, \dots, C\}$, as a polynomial of degree p

$$\mu_{ct} = \beta_{0c} + \beta_{1c}t + \dots + \beta_{cp}t^p \quad (2.2)$$

Note that the coefficients of the polynomial depend on the condition c , that is a different polynomial is allowed for each condition. Note also that a different model is fit for each gene.

A standard approach to fit the model in (2.1) is to use mixed-effect models, where the fixed effects are the true gene expressions over time across conditions and the random effects are given by the array effects. As an alternative to this, we show how differences across the conditions over time can be captured by M-quantile coefficients. In general, let $q \in [0, 1]$. The q -th conditional quantile of Y given $X = x$ is defined by

$$Q_q(x) = \operatorname{argmin}_{|a| < \infty} E[\rho_q(Y - a) | X = x],$$

where the loss function is given by $\rho_q(u) = qu$ when $u > 0$ and $\rho_q(u) = (1 - q)u$ when $u \leq 0$. It is well known that

$$q = \operatorname{Pr}[Y \leq Q_q(x) | X = x].$$

In analogy to the M-quantile extension, we can re-write this as

$$q = \frac{E[I_{Y \leq Q_q(x)} | X = x]}{E[1 | X = x]}.$$

Now we replace the L_1 loss function by the Huber proposal function $H(u)$. This function is differentiable, with the derivative $H'(u) = \psi(u)$ given by $\psi(u) =$

$u I(|u| \leq k) + k \operatorname{sgn}(u) I(|u| > k)$. The constant k depends on the level of noise and outliers in the data and is normally chosen to be 1.345. With this loss function, the conditional M-quantile is now the solution to

$$Q_q(x) = \operatorname{argmin}_{|a| < \infty} E[H(Y_a)[(1 - q)I(Y \leq a) + qI(Y > a)]].$$

After taking the derivatives, the conditional M-quantile becomes the solution to $E[\psi(Y_a)[(1 - q)I(Y \leq a) + qI(Y > a)]] = 0$. That is

$$q = \frac{E[\psi(Y - Q_q(x))I[Y \leq Q_q(x)]|X = x]}{E[\psi(Y - Q_q(x))|X = x]}. \quad (2.3)$$

This is the M-quantile coefficient, that is the quantile value of the M-quantile regression line that goes through a point x .

In our application, x is one measurement of gene expression at one of the time points so there is one M-quantile coefficient for each observed gene expression. Solving the equation (2.3) directly is complicated. As in (Chambers & Tzavidis, 2006), we derive an accurate approximation of the M-quantile coefficient, by fitting M-quantile regression lines for a large number of quantile values, using the parametric model in (2.1) and (2.2), and then using a linear interpolation to find the closest quantile value for the point in consideration.

The final aim is to detect the genes that are differentially expressed over time across the conditions. Rather than working out a test on the gene expression levels themselves, we expect to achieve greater robustness by testing for differences on the M-quantile coefficients over time. That is because these values are computed from all the quantiles of the data distribution and as such encapsulate the information from the whole distribution (and not just its mean or median). The idea is that, if a gene is expressed differently across conditions, for example it is over-expressed in one condition and under-expressed in the other, then this will be reflected in the M-quantile coefficients. That is the M-quantile coefficients for one condition will be significantly lower or higher than the M-quantile coefficients of the other condition. Of course, the temporal aspect of the data should be taken into consideration too, so the question is the one of detecting differences in the M-quantile coefficients over time. Similarly to (Vinciotti et al., 2006), though now on the M-quantile coefficients, we fit a polynomial model to the M-quantile coefficients over time and then use a Hotelling T^2 -test to detect significant differences amongst the parameters of the polynomials across the conditions. The test can compare any number of biological conditions simultaneously and the degrees of freedom of the test are derived from the number of parameters that are estimated in the model, as described in (Vinciotti et al., 2006). In the spirit of M-regression methods, which we advocate in this paper as a way to increase robustness, we use median M-regression to

fit the polynomial model to the M-quantile coefficients. This should take care of any possible outliers still present in the M-quantile coefficients values. We believe that the use of M-regression for the temporal profiles combined with the use of M-quantile coefficients to summarise the information from a large range of quantiles makes this methodology extremely robust against outliers, which are very common in microarray data as well as many other applications.

3 Simulation

We have simulated data with 2000 genes, for two biological conditions and across 9 time points. The first 1000 genes are simulated as not differentially expressed between the two conditions, with their expression drawn from a $N(0, \sigma^2)$ distribution. The remaining 1000 genes are simulated as differentially expressed, with the temporal profile linearly increasing for condition 1 and linearly decreasing for condition 2. For all genes, the noise σ was varied uniformly between 0.03 and 1.2, whereas the slope of the linear profile for the differentially expressed genes was varied uniformly between 0.005 and 0.2. Figure 1 shows the histogram of p-values obtained using the Hotelling T^2 -test to compare the temporal profiles of the M-quantile coefficients, as described in the previous section.

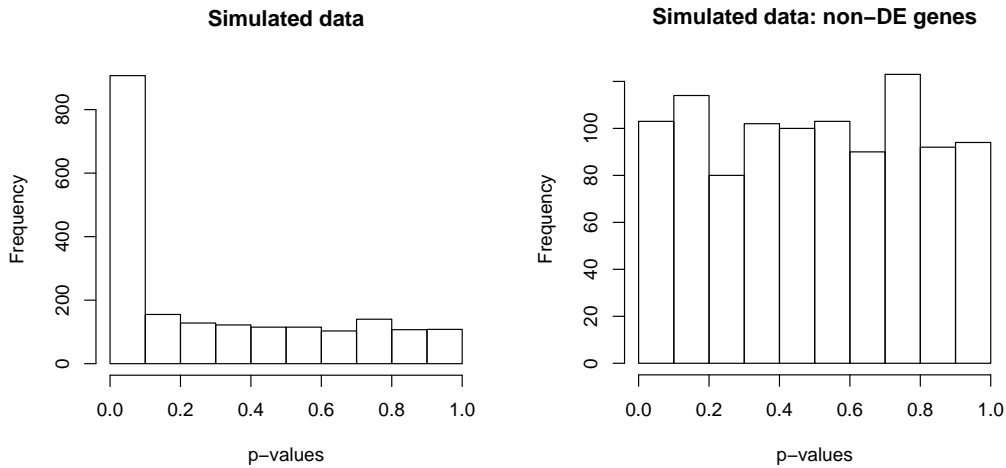


Figure 1: Histogram of p-values for all the genes (left panel) on the simulated data, with the p-values obtained from the M-quantile regression test. The right panel shows the histogram of p-values for the 1000 non-differentially expressed genes in the dataset.

The only parameter to choose is the degree of the polynomial in (2.2). For this, we choose throughout $p = 2$, that is a quadratic polynomial. This was chosen also in (Vinciotti et al., 2006) and (Yu et al., 2007), with the aim of keeping the number of parameters in the model low, given the small sample-size normally available in gene expression applications, and of making the interpretation of the results easier from a biological point of view, in terms of up-regulated and down-regulated genes. However, the methodology described in the paper is completely flexible with respect to the parametric M-quantile regression model. Figure 1 shows a good distribution of p-values for the test. The non-differentially expressed genes in the dataset show a uniform distribution of p-values (right plot), as one would expect, whereas the differentially expressed genes have all p-values close to zero.

Figure 2 shows a comparison of the M-quantile coefficient method with two more standard approaches. In the first method, we simply use a linear regression model on the gene expression data, or equivalently a mixed-effect model if array effects were also simulated (function *lm* in R). A different model is fit to each condition and the profiles are then compared using a Hotelling T^2 -test. This is the standard approach often used in the literature and is the same approach used in (Vinciotti et al., 2006). In the second approach, we use median M-regression to fit the temporal profiles on the gene expression data across conditions (function *rlm* in R). Here the aim is to increase the robustness of the method by using M-regression models directly on the data, rather than mean regression models. The third method is the one that we suggest in this paper: a two-step procedure whereby first the M-quantile coefficients are estimated over time using quadratic M-quantile regression models for values of the quantile q ranging between 0.001 and 0.999, with step 0.01, and secondly a median M-regression model is fit to the M-quantile coefficients, one for each condition. Note that the M-quantile coefficients are estimated from all the data, for all the conditions, compared to the other methods where different models are fit to each condition separately. So one further advantage of the methodology presented in this paper is that the variability in the data is better accounted for.

Figure 2 shows the ROC curves for each of the three methods, after adjusting the p-values with a Benjamini-Hochberg (BH) correction for multiple testing (function *multtest* in R). We expect the M-quantile coefficient test to perform better than the other methods, especially in the presence of noise. This is exactly what Figure 2 shows: when the noise is drawn from a normal distribution (left panel), the methods are all performing similarly well, but when the noise is contaminated (right panel) then the M-quantile coefficient test is performing best, followed by the M-regression model and finally by the mean regression model, the least robust of all. We have considered different ways of contaminating the noise, either by using a mixture of normals or a mixture of normal and exponential noise, and found that the out-performance of the method is strictly related to the degree of contamination.

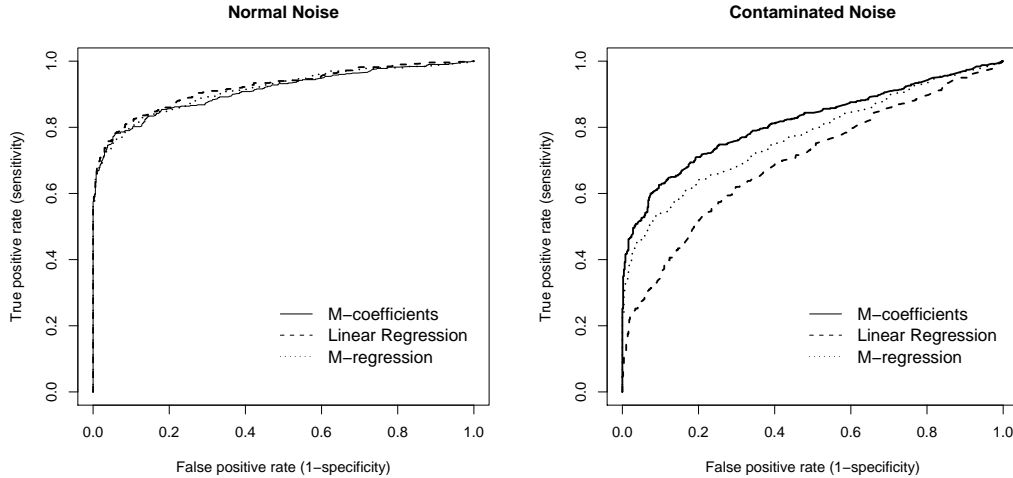


Figure 2: ROC curves comparing the M-quantile coefficient test with the more standard M-regression and mean regression tests.

The results in Figure 2 show quite a severe case where about a third of the data is contaminated by an exponential distribution with parameter $\lambda = 0.5$ and the remaining data is simulated from a normal distribution. This simulates the case when unusually large values of gene expression data are observed.

4 Application

In this section, we show a real application of this methodology, where we have applied the test based on the M-quantile coefficients on microarray data from wild-type mice and three mouse strains with different forms of muscular dystrophy. The aim is to identify genes with differences in temporal expression profiles between the strains. The four mouse strains were profiled at different ages: dystrophin-, beta-sarcoglycan and gamma-sarcoglycan deficient mice, and wild-type mice. The first three are animal models for different muscular dystrophies. The data contains 7144 genes, whose expression is measured over 9 time points. For each gene, we have available 16 time series profiles, four for each class. The data and biological problem are further described in (Vinciotti et al., 2006). In (Vinciotti et al., 2006) and (Yu et al., 2007), we analyse the same dataset based on the log-ratios of gene expression for the two channels from each array. Taking log-ratios is quite a standard technique in microarray analysis, to overcome the possibility of spot ef-

fects. By taking log-ratios, one can only estimate the gene expression profile with respect to one of the time points, normally the first one. In this paper, we model the log-intensity values for each channel separately, so that information can be obtained also at the first time point. Array effects are accounted for by including the corresponding random effects in the mixed model (Equation 2.1).

We have run the test based on the M-quantile coefficients on this dataset to detect genes with the temporal profile significantly different across the four different biological conditions. The model and the test are run independently for each of the 7444 genes and a standard multiple testing method is used to adjust the resulting p-values (function *multtest* in R). We have compared the results with a standard approach where a mixed-effect model is used to fit the gene expression profiles over time (function *lme* in R). Purely based on the number of differentially expressed genes, at the 5% cut-off, the M-quantile method detects 3559 genes as differentially expressed, whereas the standard approach detects 5364 genes. First of all, these numbers are much larger than the ones found in previous work (810 genes detected as differentially expressed with the method in Vinciotti et al. (2006)). However, the results are not directly comparable, as previous methods were based on log-ratios of gene expression and the temporal profiles were estimated with respect to the first time point. In fact, many genes are expected from biology to show significant differences right at the first time point. A simple t-test on the gene expression data at the first time point finds 2289 genes as differentially expressed between at least one pair of conditions. Secondly, as expected, the number of genes found by the M-quantile method is lower than the standard approach, as we expect the method to be more robust against noise and as such to filter out many more genes. Indeed, 85% of the genes detected by the M-quantile method are detected also by the mixed-effect method.

We have checked the six genes that were biologically validated in (Vinciotti et al., 2006) by means of qPCR experiments, namely *Dlk1*, *Dpp4*, *Tcap*, *Myoz2*, *Dbp* and *Casq2*. The new test detects *Dpp4*, *Myoz2* and *Casq2* as differentially expressed, whereas *Dlk1*, *Tcap* and *Dbp* as not-differentially expressed. Figure 3 shows the profiles for some representative genes. The left panel shows the fit of the mixed effect model, one for each condition, whereas the right panel shows the temporal profile of the M-quantile coefficients, again one for each condition. The results are quite similar to the ones found in Vinciotti et al. (2006), with genes like *Casq2* (first row) being detected by both methods. *Myoz2* (second row) and *Dlk1* (third row) are the two genes where the results differ: *Myoz2* is found differentially expressed only by the M-quantile method, whereas *Dlk1* is not found differentially expressed by the new method. Interestingly, the expression data for *Myoz2* shows quite a high variability, as pointed out also in Vinciotti et al. (2006) so it seems to be a good candidate to show the increased robustness of the method. *Dlk1* instead

shows some differential expression at the first and last time point, which is evident both at the gene expression and M-quantile coefficient profiles. However, this difference does not seem to pass the Hotelling T^2 -test for the M-quantile coefficients.

5 Discussion and Conclusion

In this paper, we describe a new method to detect differentially expressed genes from temporal profiling datasets. The aim is to find genes whose temporal expression profile is significantly different across a number of biological conditions. The presence of both the temporal dimension as well as the different biological conditions makes this task particularly interesting.

The method suggested in this paper is based on the use of M-quantile regression to model the temporal profile of a gene under a particular biological condition. Like quantile regression (Yu et al., 2007), this method does not require any strong assumption on the error distribution, and as such it differentiates itself from the vast majority of the literature on gene expression data where the expression of a gene across a number of experiments is modelled via a normal distribution. Furthermore, in contrast to standard mean regression and even quantile regression, M-quantile regression is known to be more robust against possible large outliers, by the use of specific influence functions. This advantage is of particular appeal to gene expression data, as the sample size is normally quite small and the data notoriously quite noisy.

To further increase the robustness of the method, we summarise the M-quantile regression models for a large range of quantile values into an M-quantile coefficient. By construction, the individual M-quantile coefficient of an observation (which range between 0 and 1) represents a dimensionless measure of the residual heterogeneity in the response after heterogeneity in the design has been conditioned away. This is similar to (Chambers & Tzavidis, 2006), except that now also the temporal aspect of the data is taken into consideration, deriving in the estimation of M-quantile coefficient profiles over time. Finally, we employ a Hotelling T^2 -test to detect significant differences of the temporal M-quantile profiles across a number of biological conditions. We have shown the increased robustness of the method on simulated data and an application of the method on real data from microarray experiments.

The work in this paper can be extended in different directions. Firstly, quantile and M-quantile regression methods can be extended to model interactions between genes, thus aiding the statistical reconstruction of gene regulatory networks. Secondly, the approach does not make any assumption on the error distribution and as such it can be applied to a variety of other applications.

Acknowledgement

We thank Prof Ray Chambers and Dr Nikos Tzavidis for giving us the M-quantile R code used in (Chambers & Tzavidis, 2006) and Dr Peter-Bram 't Hoen for letting us use the muscular dystrophy gene expression data.

References

- K. Archer & T. Guennel (2006). 'An application for assessing quality of RNA hybridized to Affymetrix GeneChips'. *Bioinformatics* **22**(21):2699–2701.
- M. Breckling & R. Chambers (1988). 'M-quantiles'. *Biometrika* **75**(4):761–771.
- R. Chambers & N. Tzavidis (2006). 'M-quantile models for small area estimation'. *Biometrika* **93**(2):255–268.
- L. Chen & G. Yin (2002). 'Computing the Efficiency and Tuning Constants for M-Estimation'. *ASA Proceedings of the Joint Statistical Meetings* pp. 478–482.
- A. Giloni, et al. (2006). 'Robust weighted LAD regression'. *Computational statistics and data analysis* **50**:3124–3140.
- P. Huber (1964). 'Robust estimation of location parameter'. *Annals of Mathematical Statistics* **35**:73–105.
- R. Koenker (2005). *Quantile regression*. Cambridge University Press, Cambridge.
- P. Kokic & R. Chambers (1997). 'A measure of production performance'. *Journal of Business and Economic Statistics* **15**(4):445–451.
- Y. Luan & H. Li (2003). 'Clustering of time-course gene expression data using a mixed-effects model with B-splines'. *Bioinformatics* **19**(4):474–482.
- P. Ma, et al. (2006). 'A data-driven clustering method for time course gene expression data'. *Nucleic Acids Research* **34**(4):1261–1269.
- P. Ma & W. Zhong (2008). 'Penalized clustering of large scale functional data with multiple covariates'. *Journal of the American Statistical Association* **103**:625–636.
- B. Mendes & D. Tyler (1996). 'Constrained M-estimates for regression'. In H. Rieder (ed.), *Lecture Notes in Statistics*, no. 109, pp. 299–320. Springer-Verlag.

- S. Ng, et al. (2006). ‘A mixture model with random-effects components for clustering correlated gene expression profiles’. *Bioinformatics* **22**(14):1745–1752.
- T. Park, et al. (2003). ‘Statistical tests for identifying differentially expressed genes in time-course microarray experiments’. *Bioinformatics* **19**(6):694–703.
- J. Storey, et al. (2007). ‘The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments’. *Biostatistics* **8**(2):414–432.
- J. Storey, et al. (2005). ‘Significance analysis of time course microarray experiments’. *PNAS* **102**(36):12837–12842.
- V. Vinciotti, et al. (2006). ‘Exploiting the full power of temporal gene expression profiling through a new statistical test: Application to the analysis of muscular dystrophy’. *BMC Bioinformatics* **7**(183).
- H. Wang & X. He (2007). ‘Detecting differential expressions in GeneChip microarray studies: a quantile approach.’. *Journal of the American Statistical Association* **102**:104–112.
- H. Wang & X. He (2008). ‘An enhanced quantile approach for assessing differential gene expressions’.
- Y. Wang, et al. (2007). ‘Robust estimation using the Huber function with a data-dependent tuning constant.’. *Journal of Computational and Graphical Statistics* **16**:468–481.
- S. Young & A. Bowman (1995). ‘Non-parametric analysis of covariance’. *Biometrics* **51**:920–931.
- K. Yu, et al. (2007). ‘Bayesian median regression for temporal gene expression data.’. In *Proceedings of CompLife 2007*.
- M. Yuan & C. Kendzierski (2006). ‘Hidden Markov models for microarray time course data in multiple biological conditions.’. *Journal of the American Statistical Association* **101**(476):1323–1340.

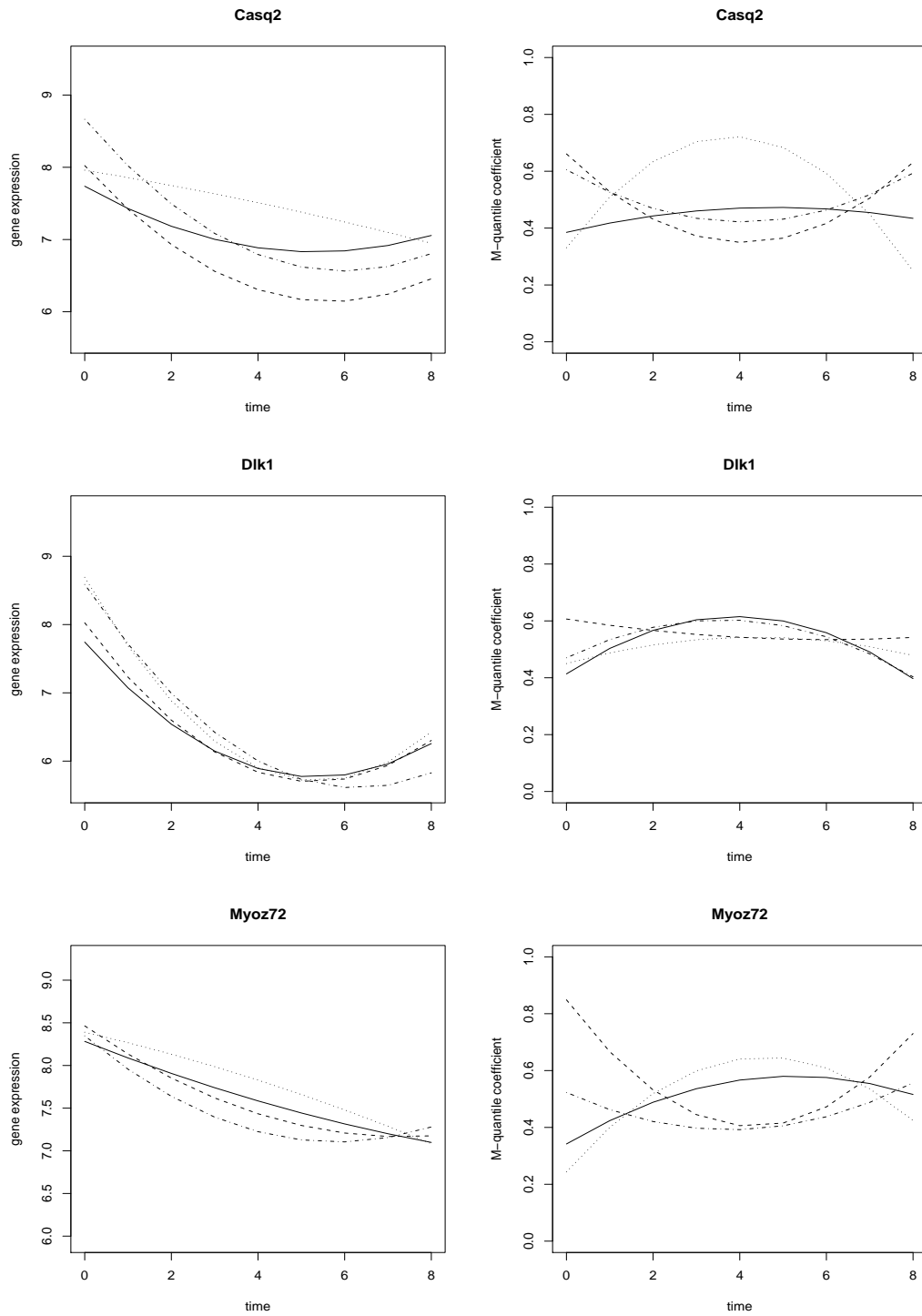


Figure 3: Mixed-effect model on expression data (left) and temporal profile of the M-quantile coefficients (right panel) for three genes previously validated using qPCR experiments. The different lines corresponding to the four different biological conditions.