Discussion Paper

Modeling mobility trends - update based on new ODiN data and level breaks

Harm Jan Boonstra, Jan van den Brakel and Sumonkanti Das

February 11, 2021

Abstract

This work is carried out by Statistics Netherlands in collaboration with KiM/Rijkswaterstaat as an extension to the trend series project carried out in 2018/2019 in which time series multilevel models have been developed for estimating mobility trends (Boonstra et al., 2019). In this extension new data from 2018 and 2019 are added to the series. These data are based on the new ODiN design of the Dutch Travel Survey. This means that the model must now include level break coefficients accounting for this latest redesign. In addition, the age classification has been enhanced to better match the desired level of detail of published trend analyses.

We describe the updated models for the two target variables considered: the number of trip legs per person per day and the distance traveled per trip leg. The models are specified in a hierarchical Bayesian framework and estimated using a Markov Chain Monte Carlo simulation method. From the model outputs smooth trend estimates can be computed at various aggregation levels for the mean number of trip legs per person per day and the mean distance traveled per trip leg, as well as for derived quantities such as the mean distance per person per day.

This report also discusses several attempts to further improve the model, in particular regarding travel modes ”Walking” and ”Cycling”.

1 Introduction

The Dutch Travel Survey (DTS) is a long-standing annual survey on mobility of residents of the Netherlands. It is carried out by Statistics Netherlands (CBS) and important users of the data are Rijkswaterstaat and The Netherlands Institute for Transport Policy Analysis (KiM, Kennisinstituut voor Mobiliteitsbeleid), both part of the Ministry of Infrastructure and Water Management.

Since 1985, the DTS survey has undergone several redesigns. The redesigns in 1999, 2004, 2010 and 2018 have caused major discontinuities in the time series of estimates on mobility. In 2004 the design actually remained largely unchanged, but its implementation was transferred to another agency, causing several changes in the observed series. For brevity, however, we will mostly also refer to this transition as a ’redesign’.

For users of mobility estimates the changes due to redesigns are very inconvenient as they hamper the temporal comparability. For the redesign of 1999 direct information was available on the sizes of the discontinuities, based on a parallel conducted pilot study. This has been used to correct the series of estimates prior to 1999 to the level of the estimates under the new design. For the redesigns of 2004, 2010 and 2018 such parallel studies have not been carried out, so in order to estimate the discontinuities a time series model is needed, see van den Brakel et al. (2017). The time series models developed in the current trend estimation project aim to account for the discontinuities due to the redesigns, such that reliable series of trend estimates are obtained with good comparability over time.

Another issue addressed in the trend estimation project is the fact that estimates are desired for a breakdown into many domains, meaning that for each domain determined

by both person characteristics (sex and age class) and trip characteristics (purpose and transportation mode) reliable time series of estimates are to be produced. So not only the discontinuities in each of these series should be accounted for, but in addition the amount of data directly relevant to an estimation domain in a specific year is often so small that direct estimates are very noisy and unreliable. The time series of such direct estimates display a lot of volatility caused by the large variances. The time series models developed are able to reduce the noise and obtain smoother series of estimates by 'borrowing strength' over time as well as over multiple domains. Here the 'borrowing of strength' over domains is brought about by using multilevel time series models with random effects for several levels defining the domains. Within the field of official statistics, the framework of using models to smooth estimates over domains of interest is known as small area estimation, see Rao and Molina (2015) for an overview.

The overall purpose of the mobility trends project has been described by the initiators of the project (KiM, Rijkswaterstaat, CBS) as 'Development of a statistical methodology that can derive reliable trend estimates from OVG-MON-OViN-ODiN sample data for the most prevalent mobility data and that deals in a robust way with discontinuities due to redesigns of the survey process and sample noise.' Here OVG, MON, OViN and ODiN refer to the various names used for the DTS during periods with different survey designs. To achieve the purpose as described, time series multilevel models are employed to fit the input data, consisting of direct estimates and estimated standard errors compiled from the DTS survey data. The resulting trend estimates are used by KiM for example in their publication 'Mobiliteitsbeeld' containing actual figures, trends, and expectations about mobility in the Netherlands. The trend estimates up to 2017 have also been published on Statistics Netherlands' publication database StatLine along with the regular annual output based on the DTS. It is anticipated that due to the recent ODiN redesign it may take a few years before the latest discontinuities can reliably be estimated and the published trends can be updated.

In Boonstra et al. (2019) the results of the trend estimation project based on mobility data from 1999 until 2017 are described. That analysis therefore does not include the most recent data based on the new ODiN design. In this report we update the results of Boonstra et al. (2019) to include the new data, accounting for the level breaks introduced by the latest redesign.

The two target variables that are modeled using time series multilevel models are:

- number of trip legs per person per day (pppd)
- distance per trip leg (in hectometers)

A trip for a certain purpose may consist of several trip legs characterized by different transportation modes. Estimates are computed for domains defined by a cross-classification of some or all of the following classification variables:

- sex (male, female)
- ageclass (0-5, 6-11, 12-17, 18-24, 25-29, 30-39, 40-49, 50-59, 60-64, 65-69, 70+)
- purpose ("Work", "Shopping", "Education", "Other")
- mode ("Car driver", "Car passenger", "Train", "BTM (bus/tram/metro)", "Cycling", "Walking", "Other")

The age classification has been extended compared to that used in Boonstra et al. (2019): the age groups 18-29 and 60-69 have now been split into respectively 18-24, 25-29 and 60-64, 65-69. The 0-5 age group is no longer observed in ODiN, but for now we keep modelling this group. This means, however, that estimates for the 0-5 age group from 2018 onwards are extrapolations and may therefore be less reliable.

The purpose category ”Other” includes family and social visits, recreation, as well as business visits. Combining these categories already reduces some discontinuities associated with the 2004 and 2010 redesigns, see e.g. Willems and van den Brakel (2015). The mode category ”Other” includes for example motorcycle, boat and skates.

The time series multilevel models are defined at the most detailed level, corresponding to the full cross-classification of sex, ageclass, purpose and mode, giving rise to $2 \times 11 \times 4 \times 7 = 616$ estimates for a particular year (up from 504 using the previous age classification), although some of them such as car-driving or working young children are structurally zero.

As observed in Boonstra et al. (2019) modeling the level break effects as fixed effects generally results in overestimated discontinuities. To reduce the risk of overestimated discontinuities and overfitting in general, we model many effects including discontinuities associated with the design transitions as random effects instead. In particular, a regularization method that employs non-normally distributed random effects is used to suppress noisy model coefficients and at the same time allow large effects sufficiently supported by the data. Outliers in the input direct estimates are also modeled, either by adopting a sampling distribution with broader than normal tails or by modeling them explicitly as additional random effects, which are subsequently removed from the trend estimates.

The remainder of this report is organized as follows. Section 2 describes the data sources used including a brief overview of the different redesigns the DTS has undergone. In Section 3 the computation of direct estimates and variance estimates from the DTS survey data is discussed, along with transformations of direct estimates and the Generalized Variance Function approach for smoothing the variance estimates, which both improve model fitting. Section 4 describes the hierarchical Bayesian time series multilevel modeling framework. The (updated) models for trip legs and distance are presented in Section 5. Section 6 provides a discussion of the trend estimates based on the estimated models, and compares with some results based on several model adjustments. Model evaluation results including a revision analysis are given in Section 7. The paper concludes with a discussion in Section 8. An appendix contains figures of the estimated trend series based on the selected models.

2 Data sources

The DTS is an annual survey that attempts to measure the travel behaviour of the Dutch population. Each year, a sample is drawn with sampling units being defined either as households (before 2010), or persons (since 2010). The variables of interest considered in this study are the number of trip legs and the distance traveled. Direct estimates for these quantities can be obtained using the survey weights that are computed for each year’s response data. The survey weights reduce the bias due to non-response, and the estimates based on them correspond to the general regression (GREG) estimator (see e.g. Särndal et al. (1992)).

The DTS started in 1978, and originally was known under the (Dutch) name Onderzoek Verplaatsingsgedrag (OVG). It started off as a face-to-face household survey where each household member of 12 years or older was asked to report his/her mobility for two days. In 1985 the first large redesign took place. Interview modes changed to telephone and postal, and respondents reported their mobility on one specific day. This redesign

led to discontinuities in the annual series of some of the statistics based on OVG. In 1994 the sample size of the DTS was substantially increased and from that year children under 12 years old have also been included in the surveyed population of interest. In 1999, the DTS went through the second major redesign that featured some response motivation and follow-up measures. In preparation to this redesign a pilot based on the new design was conducted in 1998 in parallel with the survey under the old design. Based on the parallel surveys, correction factors were computed to correct the 1985-1998 OVG to the level of the new OVG. In 2004, the data collection for the survey was transferred to another agency. The survey design remained largely unchanged except for smaller sample sizes and some methodological changes. This 2004 transition also gave rise to discontinuities in some of the series, notably those disaggregated by purpose. The DTS during the period from 2004 until the next major redesign in 2010 is referred to as MON (Mobiliteitsonderzoek Nederland). Since 2010 the DTS has been conducted by Statistics Netherlands again. In 2010 the survey changed to a person survey, and a sequential mixed mode design with face-to-face, telephone and web modes was established. This changeover led to sizeable discontinuities in many series. The years 2010 to 2017 constitute the OViN (Onderzoek Verplaatsingen in Nederland) period of the DTS. For more information on the history of the DTS and the changes made by the redesigns, we refer to Konen and Molnár (2007), Molnár (2007) and Willems and van den Brakel (2015).

Starting from 2018 another design is in place, named ODiN (Onderweg in Nederland). In ODiN several changes have been adopted, most of which can have systematic effects on the level of the observed mobility characteristics. Among the most important changes are

- The questionnaire has been completely redesigned.
- Children aged 0-5 are no longer surveyed.
- The definition of 'regular' mobility, which is the definition used in most DTS-based publications, has changed, and now includes domestic holiday and professional mobility. Flight trips are no longer observed in ODiN.
- ODiN is a CAWI-only survey, where OViN used a sequential mixed-mode CAWI-CAPI/CATI strategy. To boost response rates incentives in the form of (a chance to win) tablets are now being used.
- A new weighting scheme is used to better account for the changed composition of the response due to the design changes, including changes in regional oversampling.

The DTS considers only mobility within the Netherlands. Also, the DTS uses the concept of regular mobility, which until 2017 meant the mobility excluding holiday mobility (both domestic and abroad) and professional transportation mobility. Therefore, trips with purpose holiday and professional transportation have been removed from the survey data, as far as possible. As mentioned, however, the definition of regular mobility has changed under the ODiN redesign and now includes both domestic holiday and professional mobility. In the case of professional mobility, the data allow to identify such trips, and it has been decided here to exclude professional mobility, to be more consistent with the definition of mobility used so far in the trend estimation project. However, the ODiN questionnaire design does not easily allow to remove the observed domestic holiday mobility. This means that the ODiN discontinuities will include the effects of this change in mobility definition.

For the OViN years the data contain a small number of trips for children under the age of 12 with purpose "Work", and we have changed this to purpose "Other". Flight trips are also removed from the data, because they are no longer reported in ODiN and because

they gave rise to some unstable estimates of distance traveled for mode ”Other”. For now, data about all age classes are used. In the future it might be decided to drop the data for young children aged 0-5 years, since ODiN no longer observes this category.

Even though the DTS dates back to 1978, it was decided in Boonstra et al. (2019) to only use DTS data starting from 1999, the first year of the new OVG survey. This turned out to be sufficient for the purpose of obtaining reliable trends over the last 15 years or so. It also means that the large discontinuities arising from the 1999 redesign need not be modelled.

It is considered important that mobility trend estimates based on the DTS are in line with external data sources on mobility. Such information has been used in a plausibility analysis, and a few external sources have also been considered for use as auxiliary information in the time series models used for trend estimation. One such source is a time series of annual total passenger train kilometers based on passenger surveys run by the Dutch railways NS. These series also include data on train rides by other private companies active in the Netherlands. Another relevant data source is the annual series of road intensities, compiled by Statistics Netherlands from road induction loop data. In addition, there is a time series of registered kilometers driven by cars from Nationale Autopas (NAP). This series includes kilometers driven abroad but nevertheless it is a potential covariate in the time series models based on the DTS. Finally, annual figures for a set of weather characteristics coming from the Royal Netherlands Meteorological Institute (KNMI) have been considered as additional auxiliary series.

3 Direct estimates

Basing a time series model for mobility trends directly on the microdata from all years would require a very complex model that must account for non-response, different aggregation levels of interest, discontinuities, time trends, etc, all at once, which would be computationally intractable. Instead we follow a two-step estimation procedure often used in small area estimation. In the first step, estimates and variance estimates of the target variables are obtained directly from each year’s microdata, at the most detailed aggregation level of interest. Here we make use of the existing survey weights, accounting for the sampling design and non-response. In the second step these so-called ’direct estimates’ serve as input for a time series model, which can be used to compute smoothed estimates of mobility accounting for possible discontinuities caused by the redesigns. This section outlines the computation of the direct estimates from the OVG-MON-OViN-ODiN survey data. For additional details, see Boonstra et al. (2018).

The direct estimates are computed for all years from 1999 until 2019 for trip legs pppd and distance per trip leg. This results in two tables of 616 series of direct estimates at the most detailed breakdown level considered.

3.1 Point estimates

Point estimates are readily computed using the existing survey weights. First consider the number of trip legs, and let r_i denote the number of trip legs reported by person i for the surveyed day. The average number of trip legs pppd is then estimated by

$$\hat{R} = \frac{\sum_{i \in s} w_i f_i r_i}{\sum_{i \in s} w_i f_i}, \tag{1}$$

where the sums run over respondents, w_i are person weights satisfying $\sum_{i\in s} w_i = N$ with N the total population size, and f_i is a so-called vacation factor. The latter take values slightly less than 1, and are used to account for vacation mobility. The vacation factors are based on estimates obtained from the CVO (Continu Vakantieonderzoek) survey. They can be derived from the 'trip weights' v_i as

$$f_i = \frac{v_i}{Dw_i}, \tag{2}$$

where D is the number of days in a year. The estimates (1) can be written more compactly in terms of the trip weights as

$$\hat{R} = \frac{\sum_{i\in s} v_i r_i}{\sum_{i\in s} v_i}. \tag{3}$$

The vacation factors have been used for official publications based on OVG, MON and OViN. In ODiN the vacation factors are no longer used, since a correction for non-response bias regarding vacation mobility is now integrated into the person weights w_i by using estimated population totals from CVO in the weighting scheme directly.

For the second target variable of interest, distance, we estimate the average distance per trip leg by

$$\hat{A} = \frac{\sum_{i\in s} w_i f_i a_i}{\sum_{i\in s} w_i f_i r_i} = \frac{\sum_{i\in s} v_i a_i}{\sum_{i\in s} v_i r_i}, \tag{4}$$

where a_i is the total distance for person i for all trip legs.

For estimates by mode and/or purpose, each particular category defines specific variables r and a referring only to the trip legs in that category, so that equations (1) and (4) still apply. For (further) subdivisions with regard to the person characteristics sex and ageclass, it is convenient to introduce a dummy variable δ_i for each combination of sex and ageclass, being 1 if person i belongs to this group and 0 otherwise, and then write instead of (1) and (4),

$$\begin{aligned} \hat{R} &= \frac{\sum_{i\in s} w_i f_i \delta_i r_i}{\sum_{i\in s} w_i f_i \delta_i}, \\ \hat{A} &= \frac{\sum_{i\in s} w_i f_i \delta_i a_i}{\sum_{i\in s} w_i f_i \delta_i r_i}. \end{aligned} \tag{5}$$

By using δ_i also in the denominator of $\hat{R}$, we obtain estimates of the means per sex, ageclass combination. Note that the denominator of $\hat{R}$ does not depend on any selection of purpose or mode.

As mentioned in the Introduction, at the most detailed level, each target variable gives rise to a set of 616 estimates per year, corresponding to the full cross-classification of person characterics sex and age class and trip characteristics purpose and mode. Some of the 616 domains are, however, non-existent. We refer to these domains as structural zeros, since the number of trips in these domains is zero by definition. This concerns the following domains: age 0-5 and 6-11 in combination with mode "Car driver" or purpose "Work" and age 12-17 mode car driver before 2011. Starting from 2011 it is possible to drive a car from age 17, and this can be seen in the data. Distances per trip leg corresponding to structural zero trip legs are undefined, and therefore missing in the set of direct estimates. Other occasional zeros for trip legs and missings for distance per trip leg occur in some years for 'rare domains' such as education for the elderly. These accidental zeros and missings will be filled in by the predictions based on the time series models.

3.2 Variance estimates

For variance estimation we distinguish between person surveys (OViN, ODiN) and household surveys (OVG, MON). For the latter, the household is the unit of sampling. Observations from persons from the same household cannot be regarded as independent. For example, distances traveled by young children and their parents are often correlated, depending on purpose and mode. Variance estimates should account for the dependence between persons clustered within households.

First write estimates (1) and (4) in the general form

$$\hat{Y} = \frac{\sum_{i\in s} w_i y_i}{\sum_{i\in s} w_i z_i}\,, \tag{6}$$

which is a ratio of two population total estimates based on person weights w_i. For the average number of trip legs pppd, $y_i = f_i r_i$ and $z_i = f_i$; for the average distance per trip leg, $y_i = f_i a_i$ and $z_i = f_i r_i$.

Basic estimates of the sampling variances of $\hat{Y}$ that ignore the variation of the weights, finite population corrections and the variance of the denominator, are given by

$$v_0(\hat{Y}) = \frac{1}{(\sum_{i\in s} w_i z_i)^2}\frac{N^2}{n}S^2(y)\,, \tag{7}$$

where n is the number of respondents, $S^2(y) = \frac{1}{n-1}\sum_{i\in s}(y_i - \bar{y})^2$ is the sample variance of y, with $\bar{y} = \frac{1}{n}\sum_{i\in s} y_i$ the sample mean of y.

These variance estimates are improved by taking into account (i) the variance of the denominator, (ii) the variance inflation due to variation of the weights (Särndal et al., 1989), and (iii) the variance reducing effect of some covariates used for stratification or weighting. The variance estimates incorporating all three improvements are computed as (see e.g. Särndal et al. (1992))

$$\begin{aligned}
v(\hat{Y}) &= \frac{n}{\left(\sum_{i\in s} w_i z_i\right)^2} S^2(we)\,,\\
e_i &= e_i^y - \hat{Y} e_i^z\,,\\
e_i^y &= y_i - x_i'\hat{\beta}^y\,,\\
\hat{\beta}^y &= \left(\sum_{i\in s} x_i x_i'/u_i\right)^{-1} \sum_{i\in s} x_i y_i/u_i\,,\\
e_i^z &= z_i - x_i'\hat{\beta}^z\,,\\
\hat{\beta}^z &= \left(\sum_{i\in s} x_i x_i'/u_i\right)^{-1} \sum_{i\in s} x_i z_i/u_i\,.
\end{aligned} \tag{8}$$

Here $S^2(we)$ is the sampling variance of $w_i e_i$, where e_i are generalized residuals, defined in terms of regression residuals e_i^y for y and e_i^z for z. The regressions are based on vectors of covariates x_i and a positive variance factor u_i. For the persons survey case we use $u_i = 1$.

For the regressions defining the residuals in (8), the following covariate model is used:

$$hhsize + province + sex * ageclass + urbanisation + month + weekday + fuel$$

in which $hhsize$ is the number of persons in a household, $ageclass$ is as defined in the Introduction, $urbanisation$ is the degree of urbanisation of the residential municipality

in 5 classes, $month$ is the survey month, $weekday$ the day in the week the response refers to, and $fuel$ is the fuel type of the car used by the respondent in three classes: petrol, other or none if the respondent doesn't use a car. These covariates represent an important subset of variables that have been used for stratification and weighting of the survey data over the years.

The variance formula (8) can be used for any variables y and z in (6) so it applies to all estimates by any combination of trip characteristics purpose, mode and person characteristics sex and ageclass.

We have compared the simple variance estimates computed with (7) with the refined ones based on (8), and observed that the differences are mostly modest but not generally negligible. The most important refinement turns out to be the variance inflation due to the variation of weights. This clearly increases the variance estimates for domain estimates based on widely varying weights.

For the years before 2010, when the surveys were conducted as household surveys, the same formulas can be used, with the understanding that the unit index i refers to households. In that case y_i, z_i refer to weighted household totals, the weights w_i to the average of the person weights within a household, and x_i to household totals of the weighting covariates. The regression variances u_i are taken equal to the household size, and n in (8) becomes the number of responding households. We refer to Boonstra et al. (2018) for further details as well as for plots of the direct estimates and their standard errors for trip legs and distance at several aggregation levels. These plots show that for 'common domains' such as purpose "Work" for age classes 30-39, 40-49, standard errors are stable and rather small. For rare domains the standard errors are on average much larger, and, like the point estimates, volatile and sometimes missing. Boonstra et al. (2018) also contains a short discussion about the covariances/correlations between the direct estimates within each year. Most of these cross-sectional correlations are small, but there are some large positive and negative ones. The largest positive correlations occur between estimates for modes that are often combined in a single trip, like "Walking" and "Train", while most negative correlations occur between modes that are rarely combined such as "Car driver" and "Cycling". Furthermore, in OVG/MON years, there are some more positive correlations induced by the household clustering, for example between estimates for parents ("Car driver") and children ("Car passenger") and purpose "Shopping" or "Other". The effect of the cross-sectional correlations on the (standard errors of) the trend estimates was tested using a simple multilevel time series model and found there to be quite small. However, due to computational problems we have not been able to use the full correlation matrices of the input estimates in the selected time series models, although we expect to see only a small effect there as well.

3.3 Transformations of input series

The direct estimates and standard errors of the number of trip legs and the distances serve as input for the multilevel time series models used to obtain more smooth and robust trend series. In Boonstra et al. (2019) it was found that instead of directly modeling the direct estimates it is better to first apply a transformation to these input estimates. Using a square-root transformation for trip legs and a log-transformation for distance was seen to improve both model fits as well as the convergence of the simulation-based model fitting procedure. These transformations also reduce the dependence between direct point estimates and standard errors. After fitting the model the inverse transformation including a bias correction is applied to produce the trend

estimates, as detailed in Subsection 5.3.

Let $\hat{Y}_{it}$ denote the direct estimate for year t and domain i of the number of trip legs or distance. For the trip legs variable a square-root transformation is used: $\hat{Y}_{it}^{\text{sqrt}} \equiv \sqrt{\hat{Y}_{it}}$. A first-order Taylor linearisation yields approximated standard errors $se(\hat{Y}_{it}^{\text{sqrt}}) = se(\hat{Y}_{it})/(2\sqrt{\hat{Y}_{it}})$. Note that these standard errors are undefined for domains without observed trips (zero point estimate and standard error), but this is no problem as they are imputed using a Generalized Variance Function (GVF) smoothing model, as described below.

For the distance variable a logarithmic transformation is used: $\hat{Y}_{it}^{\log} \equiv \log \hat{Y}_{it}$, which is well-defined as all distance input estimates are positive. Standard errors for the transformed data are approximated by first-order Taylor linearisation: $se(\hat{Y}_{it}^{\log}) = se(\hat{Y}_{it})/\hat{Y}_{it}$.

3.4 Smoothing the standard errors of the direct estimates

The time series models considered regard the (transformed) direct point estimates as noisy estimates of a true underlying signal. However, the accompanying variance estimates are largely treated as fixed, given quantities by the model. As the variance estimates can be very noisy due to the detailed estimation level, it is wise to smooth them before using them in the model. That way they better reflect the uncertainty of the direct estimates. The most obvious defect of the estimated standard errors is that they are zero in case of zero or one contributing sampling unit.[1)] This is correct for the structural zero domains, but it does not reflect the actual uncertainty about the accidental zero estimates for number of trip legs. For distance, the most problematic estimates are the zero variance estimates in case of a single contributing sampling unit. If there are no contributing sampling units for a certain domain then the direct distance estimate is treated as missing.

The models considered for smoothing the variance estimates are simple regression models relating the variance estimates to a few predictors such as sample size, design effects, and point estimates. Such models are known as Generalized Variance Function (GVF) models in the literature, see Wolter (2007), Chapter 7. As in Boonstra et al. (2019) the GVF smoothing models are applied to the transformed standard errors. The predictions from the GVF models are then used as (smoothed) standard errors accompanying the transformed direct estimates as input for the time series multilevel models. In particular, this yields reasonable standard errors for domains with no observed trips.

Let $\hat{Y}_{tijkl}^{\text{tr}}$ denote either the sqrt-transformed direct estimates for trip legs or the log-transformed estimates for distance, for year t, sex i, age class j, purpose k and mode l. For both target variables we use the same GVF smoothing model

$$\log se(\hat{Y}_{tijkl}^{\text{tr}}) = \alpha + \beta \log \tilde{Y}_{tijkl}^{\text{tr}} + \gamma \log(m_{tijkl} + 1) + \delta \log(\text{deff}_{tijkl}) + \epsilon_{tijkl}\,, \quad (9)$$

where m_{tijkl} is the number of sampling units (households or persons, depending on the

[1)] This means that there is at most one unit (person or household) in a specific sex, age class domain, who reported a trip leg for a specific purpose, mode combination.

predictor	coefficient	estimate	se
1	α	-0.688	0.010
$\log \tilde{Y}^{\text{sqrt}}_{tijkl}$	β	0.942	0.003
$\log(m_{tijkl}+1)$	γ	-0.498	0.001
$\log(\text{deff}_{tijkl})$	δ	0.399	0.006

Table 3.1 Estimated coefficients of the GVF model (9) for number of trip legs.

predictor	coefficient	estimate	se
1	α	-0.957	0.020
$\log \tilde{Y}^{\log}_{tijkl}$	β	0.211	0.011
$\log(m_{tijkl}+1)$	γ	-0.345	0.002
$\log(\text{deff}_{tijkl})$	δ	0.165	0.024

Table 3.2 Estimated coefficients of the GVF model (9) for distance per trip leg.

survey year) contributing to domain (i,j,k,l) in year t, and

$$\text{deff}_{tijkl} = 1 + \frac{var(w)_{tijkl}}{\bar{w}^2_{tijkl}}, \tag{10}$$

is the design effect of the survey weights, in which the second term is the squared coefficient of variation of the weights of the contributing units to a specific year and domain.[2)] This factor accounts for the variance inflation due to the variation of the weights. Since we cannot trust the direct estimates for very small m_{tijkl}, the $\tilde{Y}^{\text{tr}}_{tijkl}$ on the right hand side of (9) are simple smoothed estimates

$$\begin{aligned} \tilde{Y}^{\text{tr}}_{tijkl} &= \lambda_{tijkl}\hat{Y}^{\text{tr}}_{tijkl} + (1-\lambda_{tijkl})\hat{Y}^{\text{tr}}_{\cdot\cdot jkl}, \\ \lambda_{tijkl} &= \frac{m_{tijkl}}{m_{tijkl}+1}, \end{aligned} \tag{11}$$

where $\hat{Y}^{\text{tr}}_{\cdot\cdot jkl}$ denotes the mean of $\hat{Y}^{\text{tr}}_{tijkl}$ over the years and sexes. For $m_{tijkl}=0$ this replaces the estimate by the mean over year and sex for the same age class, purpose and mode. For $m_{tijkl}=1$ the average of this mean and the estimate itself is used, and for large m_{tijkl} essentially the original point estimate is used.

The regression errors ϵ_{tijkl} are assumed to be independent and normally distributed with a common variance parameter σ^2. The GVF models are fitted to the positive standard errors of the transformed direct estimates. Summaries of the estimated model coefficients for trip legs and distance are given in Tables 3.1 and 3.2. The predicted (smoothed) standard errors based on the fitted models are

$$se_{\text{pred}}(\hat{Y}^{\text{tr}}_{tijkl}) = \exp\left(\hat{\alpha} + \hat{\beta}\log\tilde{Y}^{\text{tr}}_{tijkl} + \hat{\gamma}\log(m_{tijkl}+1) + \hat{\delta}\log(\text{deff}_{tijkl}) + \hat{\sigma}^2/2\right), \tag{12}$$

where $\hat{\sigma}$ is 0.11 for trip legs and 0.43 for distance. The R-squared model fit measures for both models are quite high: 0.91 for trip legs and 0.67 for distance. Note that the exponential back-transformation in (12) includes a bias correction, which in this case has only a small effect.

2) In case of 0 or 1 contributing units we have defined deff to equal 1.

4 Time series multilevel modeling

The time series multilevel models considered are extensions of the popular basic area level model proposed by Fay and Herriot (1979). The models are defined at the most detailed level, i.e. the full cross-classification of sex, ageclass, purpose, mode and year. Let us again denote by $\hat{Y}_{it}^{\text{tr}}$ the transformed direct estimates for either trip legs or distance in year t and domain i. Here domain i refers to a particular combination of sex, ageclass, purpose and mode, so that i runs from 1 to $M_d = 616$ and t from 1 to T corresponding to the years 1999 to 2019. We further combine these estimates into a vector $\hat{Y} = (\hat{Y}_{11}, \dots \hat{Y}_{M_d1}, \dots \hat{Y}_{1T}, \dots \hat{Y}_{M_dT})'$. Note that $\hat{Y}$ is a vector of dimension $M = M_dT$. Structural zero domains are not modeled, and it is implicitly understood that they are removed from all expressions. Also removed are all age class 0-5 estimates after 2017, because this age group is no longer covered by the ODiN survey. The number of modeled initial estimates is thereby reduced from $M = M_dT = 616 \times 21 = 12936$ to a total of 11928. For distance per trip leg there are some additional domains without initial estimates due to the (coincidental) absence of observed trips. The total number of available initial distance estimates is 11337. For both target variables model estimates are eventually produced for all 11928 non-structurally-zero domains.

4.1 Model structure

The multilevel models considered take the general linear additive form

$$\hat{Y}^{\text{tr}} = X\beta + \sum_{\alpha} Z^{(\alpha)} v^{(\alpha)} + e\,, \tag{13}$$

where X is a $M \times p$ design matrix for a p-vector of fixed effects β, and the $Z^{(\alpha)}$ are $M \times q^{(\alpha)}$ design matrices for $q^{(\alpha)}$-dimensional random effect vectors $v^{(\alpha)}$. Here the sum over α runs over several possible random effect terms at different levels, such as transportation mode and purpose smooth trends, white noise at the most detailed level of the M domains, etc. This is explained in more detail below. The sampling errors $e = (e_{11}, \dots, e_{M_d1}, \dots e_{M_dT})'$ are taken to be normally distributed as

$$e \sim N(0, \Sigma) \tag{14}$$

where $\Sigma = \text{diag}\left(se_{\text{pred}}(\hat{Y}_{tijkl}^{\text{tr}})^2\right)$, i.e. a diagonal matrix with values equal to the square of the smoothed standard errors computed as discussed in Subsection 3.4.

Equations (13) and (14) define the likelihood function

$$p\left(\hat{Y}^{\text{tr}} | \eta, \Sigma\right) = N\left(\hat{Y}^{\text{tr}} | \eta, \Sigma\right)\,, \tag{15}$$

where $\eta = X\beta + \sum_{\alpha} Z^{(\alpha)} v^{(\alpha)}$ is called the linear predictor. A Student-t distribution for the sampling errors in (14) has been considered instead of the normal distribution to give smaller weight to more outlying observations. This is a traditional approach for handling outliers in Bayesian regression, see e.g. West (1984). We allow the degrees of freedom parameter of the Student-t distribution to be inferred from the data. It has been assigned a Gamma$(2, 0.1)$ prior distribution, which was recommended as a default prior in Juárez and Steel (2010).

The fixed effect part of η contains an intercept and main effects and possibly the second-order interactions for linear trends, discontinuities and the breakdown variables sex, age, purpose and mode. The vector β of fixed effects is assigned a normal prior

$p(\beta) = N(0, 100I)$, which is very weakly informative as a standard error of 10 is very large relative to the scales of the transformed direct estimates and the covariates used.

The second term on the right hand side of (13) consists of a sum of contributions to the linear predictor by random effects or varying coefficient terms. The random effect vectors $v^{(\alpha)}$ for different α are assumed to be independent, but the components within a vector $v^{(\alpha)}$ are possibly correlated to accommodate temporal or cross-sectional correlation. To describe the general model for each vector $v^{(\alpha)}$ of random effects, we suppress superscript α in what follows.

Each random effects vector v is assumed to be distributed as

$$v \sim N(0, A \otimes V)\,, \tag{16}$$

where V and A are $d \times d$ and $l \times l$ covariance matrices, respectively, and $A \otimes V$ denotes the Kronecker product of A with V. The total length of v is $q = dl$, and these coefficients may be thought of as corresponding to d effects allowed to vary over l levels of a factor variable, e.g. purpose effects ($d = 4$) varying over time ($l = 21$ years). The covariance matrix A describes the covariance structure among the levels of the factor variable, and is assumed to be known. Instead of covariance matrices, precision matrices $Q_A = A^{-1}$ are actually used, because of computational efficiency (Rue and Held, 2005). The covariance matrix V for the d varying effects is parameterized in one of three different ways:

- an unstructured, i.e. fully parameterized covariance matrix
- a diagonal matrix with unequal diagonal elements
- a diagonal matrix with equal diagonal elements

The following priors are used for the parameters in the covariance matrix V:

- In the case of an unstructured covariance matrix the scaled-inverse Wishart prior is used as proposed in O'Malley and Zaslavsky (2008) and recommended by Gelman and Hill (2007).
- In the case of a diagonal matrix with equal or unequal diagonal elements, half-Cauchy priors are used for the standard deviations. Gelman (2006) demonstrates that these priors are better default priors than the more common inverse gamma priors for the variances.

The following random effect structures are considered in the model selection procedure:

- Random intercepts for the M_d domains obtained by the full cross clasification of age, gender, purpose and mode. In this case $A = I_{M_d}$ and V is a scalar variance parameter, and the corresponding design matrix is the $M \times M_d$ indicator matrix for domains. This can be extended to a vector of random domain intercepts, random slopes for linear time effects and discontinuities due to the redesigns in 2004, 2010 and 2018. In that case V is a 5×5 covariance matrix, parameterized by variance parameters for the intercepts, linear time slopes and the coeffcients for the level interventions, and possibly 10 correlation parameters.
- Random effects that account for outliers. The data for some years appear to be of lesser quality. This is the case especially for data on the number of trip legs in 2009. In order to deal with such less reliable estimates, random effects can be used to absorb some of the larger deviations in such years. The corresponding effects are removed from the trend prediction. This is an alternative to the use of fat-tailed sampling distributions such as the Student-t distribution for dealing with outliers.
- Random walks or smooth trends at aggregated domain levels (e.g. purpose by mode). See Rue and Held (2005) for the specification of the precision matrix Q_A for

first and more smooth second order random walks. A full covariance matrix for the trend innovations can be considered to allow for cross-sectional besides temporal correlations, or a diagonal matrix with equal or different variance parameters to allow for temporal correlations only.

- White noise. In order to allow for random unexplained variation, white noise at the most detailed domain-by-year level can be included. In this case $A = I_M$ and V a scalar variance parameter, and the design matrix is $Z = I_M$.

We also investigate generalisations of (16) to non-normal distributions of random effects. Relevant references are Carter and Kohn (1996) in the state space modeling context, Datta and Lahiri (1995), Fabrizi and Trivisano (2010) and Tang et al. (2018) in the small area estimation context, and Lang et al. (2002) and Brezger et al. (2007) in the context of more general structured additive regression models. In particular, the following distributions are considered for various random effect terms:

- Student-t-distributed random effects
- Random effects with a so-called horseshoe prior (Carvalho et al., 2010).
- Random effects distributed according to the Laplace distribution. This corresponds to a Bayesian version of the popular lasso shrinkage, see (Tibshirani, 1996; Park and Casella, 2008).

These alternative distributions have fatter tails allowing for occasional large effects. The Laplace and particularly the horseshoe distribution have the additional property that they shrink noisy effects more strongly towards zero.

4.2 Model estimation

The models are fitted using Markov Chain Monte Carlo (MCMC) sampling, in particular the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). See Boonstra and van den Brakel (2018) for a specification of the full conditional distributions. The models are run in R (R Core Team, 2015) using package `mcmcsae` (Boonstra, 2020). The Gibbs sampler is run in parallel for three independent chains with randomly generated starting values. In the model building stage 1000 iterations are used, in addition to a 'burn-in' period of 250 iterations. This was sufficient for reasonably stable Monte Carlo estimates of the model parameters and trend predictions. For the selected models we use a longer run of 1000 burn-in plus 10000 iterations of which the draws of every fifth iteration are stored. This leaves $3 * 2000 = 6000$ draws to compute estimates and standard errors. The convergence of the MCMC simulation is assessed using trace and autocorrelation plots as well as the Gelman-Rubin potential scale reduction factor (Gelman and Rubin, 1992), which diagnoses the mixing of the chains. For the longer simulation of the selected models all model parameters and model predictions have potential scale reduction factors below 1.05 and sufficient effective numbers of independent draws.

Many models of the form (13) have been fitted to the data. For the comparison of models using the same input data we use the Widely Applicable Information Criterion or Watanabe-Akaike Information Criterion (WAIC) (Watanabe, 2010, 2013) and the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002). We also compare the models graphically by their model fits and trend predictions at various aggregation levels.

5 Model building, selected models, and model prediction

The transformed direct estimates for all 616 domains over the 1999-2019 period along with their smoothed standard errors serve as input data for the time series models. The variables defining the domains and the years have been used in the model development in many ways, e.g. using different interactions of various orders. Some additional covariates have been constructed in order to model the discontinuities between 2003 and 2004, between 2009 and 2010, and between 2017 and 2018, as well as to reduce the influence of some lesser quality 2009 input estimates. For the MON level breaks in 2004, a variable br_mon is introduced taking values 1 between 2004 and 2009, and 0 otherwise. For the OViN level breaks in 2010 a variable br_ovin is defined, taking values 1 for the years 2010 and later and 0 otherwise. Likewise, for the ODiN level breaks in 2018 a variable br_odin is defined, taking values 1 from 2018 and 0 otherwise. A slight modification of the break variables was necessary in order not to introduce artificial level breaks in the age 12-17 car driver domains, which are structurally zero domains before 2011. Also, for year 2009 a dummy variable $dummy\_2009$ has been created being 1 only for year 2009 and 0 otherwise. The year variable is also used quantitatively to define linear time trends, and for that purpose we use a scaled and centered version denoted $yr.c$. These variables have been used in different interactions, as fixed or random effects in the model.

Some other covariates extracted from other sources like Statistics Netherlands' Statline and KNMI meteorological annual reports have also been used as candidate covariates in the model development. The weather variables considered concern annual averages at a central measurement location in The Netherlands (De Bilt). Eventually, a variable $snowdays$ representing the number of snow days by year is used in the selected trip-legs model. In Boonstra et al. (2019) an administrative variable km_NAP representing annual registered car kilometers collected from Nationale Autopas (NAP) was used in the distance model. As this variable was not yet available for 2019, and the effect this variable has on the estimated series is only weak, we decided to remove this variable from the models considered in this report.

In the following two sub-sections, time series models developed for the number of trip legs and the distance per trip leg are discussed. Following that, it is described how the target trend estimates are derived from the developed time series models. The models are expressed as time series multilevel models in a hierarchical Bayesian framework and fit using a Markov Chain Monte Carlo (MCMC) simulation method, as described in Section 4.

5.1 Time series multilevel model for the number of trip legs

As described in Section 3, we model the square-root-transformed direct estimates of the number of trip legs pppd, using the corresponding transformed and GVF-smoothed standard errors to define the variance matrix Σ of the sampling errors.

The model parameters in (13) are separated in fixed and random effects. The fixed effects specification for the updated model is

$$sex * ageclass + purpose * mode + mode * snowdays + (purpose\_other + mode\_other + mode\_cycling) : br\_odin \quad (17)$$

The last term involving fixed effects for specific ODiN breaks has been added as compared to the model described in Boonstra et al. (2019). Besides, we recall that the variable $ageclass$ in the model specification now has more classes.

The updated random effects part of the model for number of trip legs is given in Table 5.1. The only difference compared to the previous version (Boonstra et al., 2019) is the addition of br_odin to the 'V_BR' component. This means that, apart from the additional fixed effects break terms, the ODiN breaks are modeled in the same way as the OViN and MON breaks.

Model Component	Formula V	Variance Structure	Factor A	Prior	Number of Effects
V_2009	$dummy\_2009$	scalar	$sex * ageclass * purpose * mode$	horseshoe	616
V_BR	$1 + yr.c + br\_mon\_SO + br\_ovin + br\_odin$	unstructured	$sex * ageclass * purpose * mode$	Laplace	3080
RW2AMM	$ageclass * purpose * mode$	scalar	RW2(yr)	normal	6468
RW2MM	$purpose * mode$	diagonal	RW2(yr)	normal	588
WN	1	scalar	$sex * ageclass * purpose * mode * yr$	normal	12936

Table 5.1 Summary of the random effect components for the selected time series multilevel model for trip legs. The second and third columns refer to the varying effects with covariance matrix V in (16), whereas the fourth and fifth columns refer to the factor variable associated with A in (16). The last column contains the total number of random effects for each term.

The model term named 'V_2009' in Table 5.1 is included to account for some very influential outliers in 2009. It uses a horseshoe prior distribution, which turned out to work well as for most domains the outlier effects are negligible, but for a few domains, notably those for young children and motives "Shopping" and "Other", they are very large.

The random effects component 'V_BR' includes MON, OViN and now also ODiN level break random effects, besides random intercepts, and random linear time trends, varying over all domains (the cross-classification of sex, ageclass, purpose and mode). The full covariance matix V in (16) is a 5 x 5 matrix parameterised in terms of standard deviation and correlation parameters. The variable br_mon_SO indicates that MON breaks are only considered for purposes "Shopping" and "Other". MON break effects for the purposes "Work" and "Education" resulted in somewhat artificial and implausible trend estimates and so these have been excluded from the model. This component uses a Laplace prior distribution for the random effects.

Two smooth time trend components at purpose × mode and ageclass × purpose × mode aggregation levels are included in the final model. These terms are named 'RW2MM' and 'RW2AMM' respectively in Table 5.1. For the 'RW2MM' component different variance parameters corresponding to different degrees of smoothness are allowed for all purpose, mode combinations. The different values found for the variance components of the 'RW2MM' components indeed show large diifferences in degrees of smoothness. The 'RW2AMM' component only uses a single ('scalar') variance

parameter. and can be interpreted as a correction to the 'RW2MM' trends, allowing for some differences between age classes. The contribution of the 'RW2AMM' effects is much smaller than that of the 'RW2MM' effects.

Finally, a white noise term named 'WN' in Table 5.1 captures unstructured variation over all levels of all factors. This 'WN' component accounts for more or less random variation of the true average number of trip legs pppd over the domains and the years.

5.2 Time series multilevel model for the distance per trip leg

For distance we model the log-transformed direct estimates of distance per trip leg, using the corresponding transformed and GVF-smoothed standard errors discussed in Section 3 to define the variance matrix Σ of the sampling errors. The use of Student-t distributed sampling errors in this case succeeds in reducing the influence of outliers sufficiently. The degrees of freedom parameter of the Student-t distribution is assigned a weakly informative prior and is inferred from the data.

Similar to the model for number of trip legs pppd, only main effects and second order interaction effects are used in the fixed effects part of the selected model. In the selected model the following fixed effects components are included:

$$sex * ageclass + purpose * mode + yr.c * mode + \\ + mode\_walking : br\_ovin + br\_odin \tag{18}$$

The only differences compared to the selected model in Boonstra et al. (2019) are the addition of the global ODiN break effect, and the removal of a fixed effect for kilometers driven by registered cars, for which the 2019 figure was not available in time. The term $yr.c * mode$ represents linear time trends by mode. The variables $mode\_cardriver$ and $mode\_walking$ are indicator variables for modes "Car driver" and "Walking". The term $mode\_walking : br\_ovin$ represents a single OViN break fixed effect for mode "Walking", and is included as the overall discontinuity for "Walking" appears to be too large to model using random effects only.

The other effects, including higher order interactions, are modeled as random effects, and the selected terms are shown in Table 5.2. Here the only difference with respect to the model in Boonstra et al. (2019) is the extension of the level break component 'V_BR' to accommodate ODiN break effects in the same way as discussed for the model for trip legs.

The 'V_BR' and 'WN' components are the same as in the model for trip legs. Otherwise the distance model is kept more parsimonious to prevent overfitting the more noisy input series. In particular, the model includes a smooth trend, i.e. second order random walk, for each mode, purpose combination with a variance parameter depending only on mode. This term is labeled 'RW2M' in Table 5.2.

5.3 Trend estimation and derived estimates

The trend estimates of main interest are computed based on the MCMC simulation results as follows. First, simulation vectors of model linear predictions are formed, i.e.

$$\eta^{(r)} = X\beta^{(r)} + \sum_{\alpha} Z^{(\alpha)} v^{(\alpha,r)}, \tag{19}$$

where superscript r indexes the retained MCMC draws, and each $\eta^{(r)}$ is of dimension M. Consequently, the level break effects are removed or added, depending on the choice of benchmark level. For now we choose the OViN level as the benchmark level, as in

Model Component	Formula **V**	Variance Structure	Factor **A**	Prior	Number of Effects
V_BR	$1 + yr.c + br\_mon + br\_ovin + br\_odin$	unstructured	$sex * ageclass * purpose * mode$	Laplace	3080
RW2M	$mode$	diagonal	motive * RW2(yr)	normal	588
WN	1	scalar	$sex * ageclass * purpose * mode * yr$	normal	12936

Table 5.2 Summary of the random effect components for the selected time series multilevel model for distance. The second and third columns refer to the varying effects with covariance matrix V in (16), whereas the fourth and fifth columns refer to the factor variable associated with A in (16). The last column contains the total number of random effects for each term.

Boonstra et al. (2019), even though it no longer corresponds to the most recent survey design period considered. Given the way the level break dummies are coded, it means that we need to add all OViN break effects to the predictions referring to the OVG, MON and ODiN years, and in addition need to remove the MON and ODiN effects from the predictions referring to the MON and ODiN years, respectively. Also, if present, the dummy effects for outliers are removed. We note that the survey errors e in (13) are already absent from the linear predictor (19). The simulation vectors of linear predictors thus obtained are

$$\tilde{\eta}^{(r)} = \tilde{X}\beta^{(r)} + \sum_{\alpha} \tilde{Z}^{(\alpha)} v^{(\alpha,r)}, \tag{20}$$

where $\tilde{X}$ and $\tilde{Z}^{(\alpha)}$ are modified design matrices that accomplish the stated correction for level breaks and possibly outlier effects. Back-transformation of these vectors to the original scale yields the MCMC approximation to the posterior distribution of the trends. For the square root transformation as used for modeling the number of trip legs pppd, the back-transformation amounts to

$$\theta^{(r)} = (\tilde{\eta}^{(r)})^2 + (se(\hat{Y}_{it}^{\text{sqrt}}))^2, \tag{21}$$

The second term on the right hand side accomplishes a (relatively small) bias correction using the transformed and smoothed standard errors for number of trip legs. The bias correction stems from the fact that the design expectation of the direct estimates can be written as

$$E(\hat{Y}) = E((\hat{Y}^{\text{sqrt}})^2) = E((\eta + e^{\text{sqrt}})^2) = \eta^2 + 2\eta E(e^{\text{sqrt}}) + E((e^{\text{sqrt}})^2) = \eta^2 + var(e^{\text{sqrt}}), \tag{22}$$

where e^{sqrt} is the vector of sampling errors after transformation, assumed to be normally distributed with standard errors $se(\hat{Y}_{it}^{\text{sqrt}})$.

For the log transformation, as used in modeling distance per trip leg, back-transforming $\tilde{\eta}^{(r)}$ to the original scale yields the MCMC approximation to the posterior distribution of the distance trends. The exponential back-transformation including bias correction is

$$\theta^{(r)} = e^{\tilde{\eta}^{(r)} + \text{se}(\hat{Y}^{\log})^2/2}. \tag{23}$$

The bias correction is added to largely correct a small negative bias induced by the log transformation, see for example Fabrizi et al. (2018).

The means over the MCMC draws $\theta^{(r)}$ are used as trend estimates, whereas the standard deviations over the draws serve as standard error estimates.

Recall that η and θ are vector quantities with components for all year-domain combinations. We have computed the trends at the most detailed level, but also for aggregates over several combinations of the domain characteristics. Aggregation of distance per trip leg involves the number of trip legs, and so requires combining the MCMC output for both target variables. By multiplying the distance per trip leg results by the number of trip leg pppd results we obtain the results for distance pppd. Aggregation amounts to simple summation over trip characteristics purpose and mode, and to population weighted averaging over person characteristics sex and ageclass. Inference for other derived quantities like total number of trip legs per day and total distance per day at different aggregation levels can also be readily conducted using the simulation results for the two modeled target variables.

6 Results

The appendices contain a rather complete set of time series plots for the number of trip legs pppd (A.1), distance per trip leg (A.2), and distance pppd (A.3) at different aggregation levels, including the most detailed level, based on the selected models for the number of trip legs pppd and the distance per trip leg described in Section 5. The black lines in these figures correspond to the series of direct estimates, the red lines to the model fit based on all model components, i.e. the back-transformation of (19), and the green lines to the trend series (21) or (23). In the following sub-sections, the trend estimates obtained for number of trip legs and distance are discussed, including a comparison with estimates based on some alternative model specifications.

6.1 Trip legs

Trend estimates of number of trip legs per person per day are given in Appendix A.1.

The importance of the additional fixed effects for ODiN level breaks included in (17) can be seen by comparing to the results based on the model without these fixed effects, in which ODiN level breaks are entirely modeled as random effects. These results are shown in Figures 6.1 - 6.3 for the highest aggregation levels. It is clear that without the additional fixed effects the ODiN breaks at these aggregation levels are most likely underestimated, resulting in some large and less plausible trend jumps compared to the model including these fixed effects.

As noticed in Boonstra et al. (2019), the plots in the Appendix show that the estimated trends for trip legs are hardly affected by any MON breaks. This has partly been enforced by excluding MON breaks for the purposes ”Work” and ”Education” in the random effects model component ’V_BR’. Since the OVG and MON designs are largely the same except that they were carried out by different data collection organisations, no large MON breaks were anticipated. However, at a more detailed level, some MON discontinuities are clearly present for purposes ”Shopping” and ”Other”, as shown by Figure A.6, for the 0-5 age group. It hints at a possible exchange between purposes ”Shopping” and ”Other” for young children.

By contrast, some OViN breaks are quite large. Overall, the OViN level for number of trip legs is lower than the levels observed during MON, OVG and also ODiN design periods, as is clear from Figure A.1. This is the case at least for modes ”Car driver”, ”Train” and ”Cycling”, as illustrated by Figure A.3. For mode ”Walking” there is also a large jump

Figure 6.1 Comparison of total trip leg pppd estimates between the selected model (left) and the same model with fixed ODiN break effects removed (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure 6.2 Comparison of total trip leg pppd estimates by purpose between the selected model (left) and the same model with fixed ODiN break effects removed (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

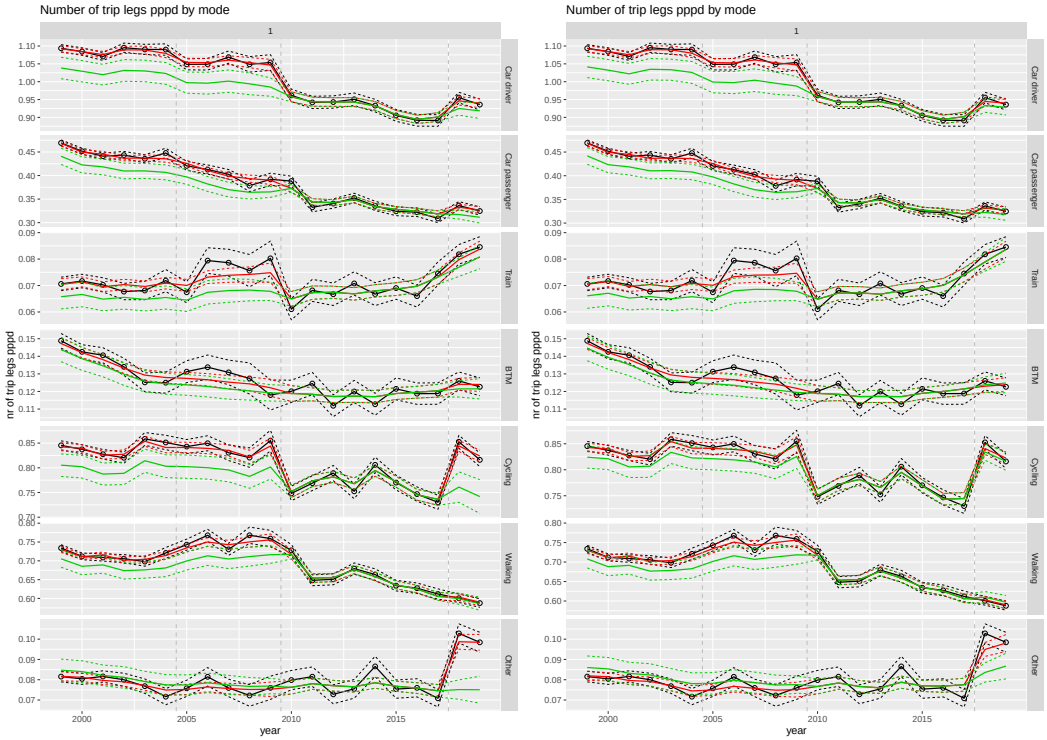

Figure 6.3 Comparison of total trip leg pppd estimates by mode between the selected model (left) and the same model with fixed ODiN break effects removed (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

downwards, but it happens in 2011, a year after the start of OViN. A full explanation of this jump is lacking, but it seems that the weather at least plays a partial role. The year 2010 was a year with a rather extreme amount of snow days, and under such circumstances it is expected that more walking trips are made, e.g. as an alternative to cycling. If the weather indeed plays a significant role in explaining the high 2010 estimate, it might be that the overall OViN discontinuity for "Walking" is underestimated, as the trend level before 2010 is not much lower than that of 2010. The small OViN break here is actually due to inclusion of $snowdays$ as a covariate. Without it, the break would have been estimated even smaller resulting in an even larger jump in 2011. In the next subsection this issue as well as the related issue of volatile trends is discussed further.

There are some noteworthy differences in discontinuities between men and women trend lines, particularly for 30-39 and 40-49 age groups, see e.g. Figure A.11. In these particular cases the differences in the levels of the direct estimates between men and women are much larger during the OViN period. As these differences are most probably due to measurement errors in OViN data, this is a drawback of basing the trend estimates on the OViN level. The particular OViN measurement errors for purpose "Education", female, age groups 30-39 and 40-49 are perhaps due to incorrect purpose assignment for mothers taking their children to school. The ODiN level for these domains seems to revert to approximately the OVG-MON level. Relatedly, an opposite movement can be discerned for purpose "Other".

Since the trends are currently defined at the level of OViN, the outcomes during the MON, OVG and ODiN periods are corrected for the discontinuities induced by the redesigns. It implies that due to the uncertainty of the estimated discontinuities the standard errors for the trend estimates in these periods are larger compared to the OViN period. This effect is strongest for the ODiN period as for now the ODiN break estimates are based on just two years of ODiN data. At an aggregated level, the standard errors of the trend estimates are even larger than the variances of the direct estimates, which only measure sampling variation. See for example Figure A.1 for estimates at the overall level and Figures A.2 and A.3 for estimates by purpose and mode.

The 2009 outlier effects are pronounced for some domains, notably for young children and purposes "Shopping" and "Other", as illustrated in Figure A.6. There is a clear exchange between both purposes for the young children. These effects have been captured by the random effect term 'V_2009' of the selected model for trip legs. The trend lines show that these outliers are indeed neutralized by excluding the 'V_2009' effects.

Tables 6.1 and 6.2 list the posterior means and standard errors of several variance components of the trip leg model. It is to be noted from Table 6.1 that the random intercepts and OViN breaks are quite strongly negatively correlated over the domains. There is also a large but more uncertain negative correlation between the random MON and ODiN break effects. Table 6.2 shows that the scales of the second order random walks by purpose and mode in the 'RW2MM' model component are very diverse. The largest scales are seen for cycling and walking for purpose "Other". This difference in volatility by purpose and mode is also visible in the trends, as shown in Figure A.4. The high volatility of the series for cycling and purpose "Other" may be a real phenomenon caused e.g. by weather effects. These series closely follow the direct estimates due to the relatively small standard errors of the latter. Other domains have more smooth trends despite volatile direct estimates, as for example "Train" for purpose "Other". As

the direct standard error estimates are much larger in this case, the model chooses a smoother trend series. The estimated standard deviation of the innovations of the smooth trend for ”Car driver”, purpose ”Work”, have increased compared to the estimated trend model described in Boonstra et al. (2019). This seems to be due to confounding with ODiN break effects, which are very uncertain given only two years of ODiN data.

Compared to the fixed effects, ’V_BR’ and ’RW2MM’ components, it was found that the more detailed trend component ’RW2AMM’ has a smaller contribution to the linear predictor, and is smoother than most ’RW2MM’ components. The white noise term ’WN’ was also found to have only a modest contribution to the trend estimates.

A positive trend in the number of trip legs can be observed for the 12-17 age group as car driver after the 2011 change in law, see Figure A.19. This domain is quite small regarding number of observed trips, but the fitted model seems to work satisfactorily here, though ODiN discontinuities are hard to estimate accurately in this case of increasing trends.

	Intercept	yr.c	br_mon_SO	br_ovin	br_odin
Intercept	13.9 (0.6)	-14.6 (6.2)	23.0 (9.1)	-46.3 (5.1)	-16.5 (7.4)
yr.c		2.0 (0.1)	17.2 (10.3)	-12.2 (7.5)	-20.9 (8.4)
br_mon_SO			1.8 (0.2)	3.0 (11.0)	-38.7 (19.2)
br_ovin				3.3 (0.2)	2.3 (10.0)
br_odin					1.8 (0.1)

Table 6.1 Estimated standard deviations and correlations ($\times 100$) for the ’V_BR’ component

	Car Driver	Car Passenger	Train	BTM	Cycling	Walking	Other
Work	10.0 (3.3)	2.1 (3.4)	2.9 (1.6)	1.8 (1.3)	2.3 (2.9)	5.3 (2.6)	2.3 (1.8)
Shopping	4.3 (2.0)	3.6 (1.9)	1.6 (1.1)	1.6 (1.8)	2.5 (1.7)	4.0 (3.1)	1.4 (1.2)
Education	1.0 (0.9)	0.7 (1.2)	0.8 (0.6)	1.9 (1.4)	8.1 (4.2)	3.5 (1.9)	2.2 (1.3)
Other	9.6 (3.2)	5.0 (2.6)	1.1 (1.0)	1.9 (1.0)	28.7 (6.9)	14.2 (4.0)	1.2 (1.3)

Table 6.2 Estimated standard deviations ($\times 1000$) with their standard errors in parentheses under the diagonal covariance matrix for the ’RW2MM’ component

6.2 Model extensions for number of trip legs

A few aspects of the model for trip legs developed in Boonstra et al. (2019) have been identified as less than optimal. These issues are examined here.

One issue relates to the fact that the trends for cycling and walking, especially for purpose ”Other”, are quite volatile. It is likely that this is the case for the underlying real unknown series as well. For one thing, the behaviour of these series is influenced by such external factors as weather conditions. We are considering only annual averages, so much of the variation in these external factors averages out, but it is plausible that the remaining annual variation can still affect the trend series.

The selected model discussed in Section 5 accounts for the dynamics of the cycling and walking series mainly through smooth second-order random walk components, labeled 'RW2AMM' and 'RW2MM' in Table 5.2. The contribution of 'RW2AMM' to the trends of cycling and walking is of a much smaller size, and is smooth. By contrast, the contribution of the 'RW2MM' component to the trends for cycling and walking is much larger, and it doesn't look smooth at all, as shown by Figure 6.4. This figure shows the contribution of the 'RW2MM' component to the linear predictor aggregated over sex and age classes. The same strong volatility is also apparent from Table 6.2 showing large standard deviations for the slope innovations for the trend series of walking and cycling, especially regarding purpose "Other". So the choice of a smooth model component seems less suitable for these two modes in combination with purpose "Other". The question arises whether the model can be improved regarding these particular trend components.

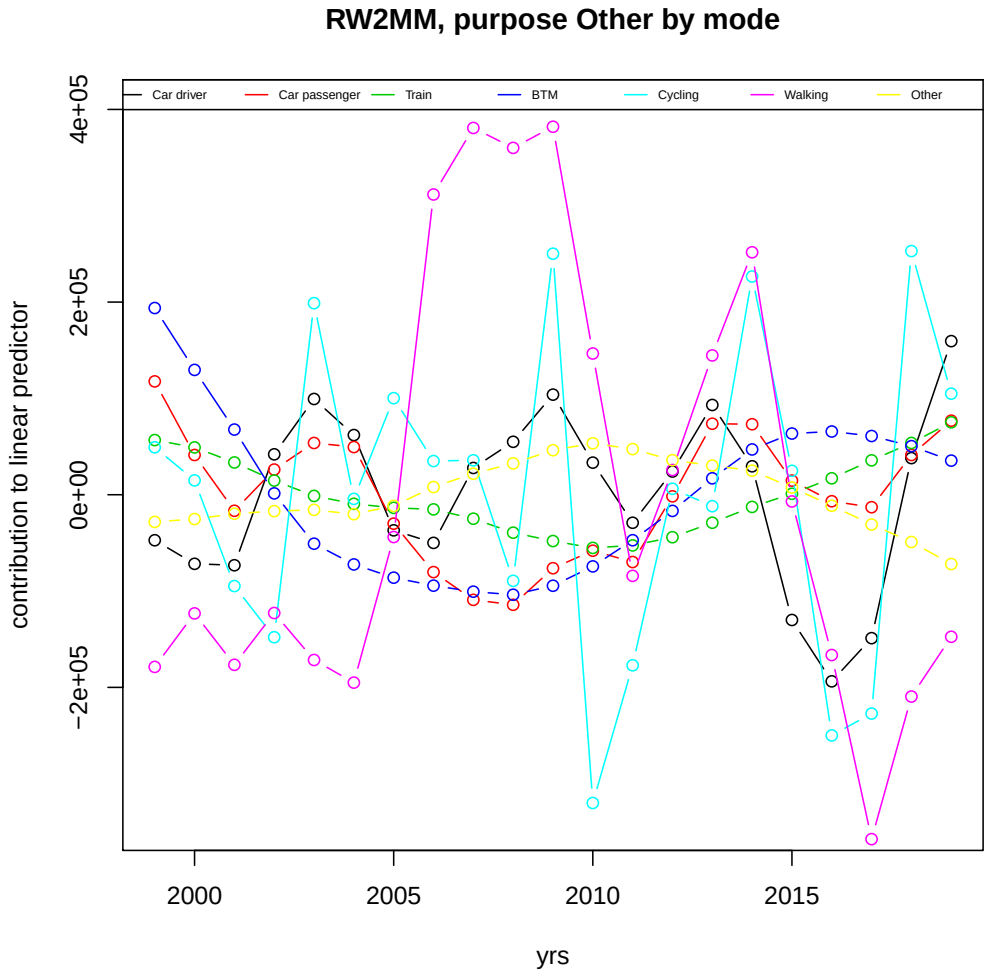

Figure 6.4 Contribution of 'RW2MM' component to the linear predictor of the trip leg model, aggregated over sex and age classes.

Following is a list of model extensions that have been attempted to address the issues with cycling and walking trends, and the main findings. In all cases the model described in Section 5.1 is used as starting point.

1. Add two first order random walks, one for walking and one for cycling, both in combination with purpose "Other". They are allowed to be correlated. The results

from this model extension show signs of overspecification: the existing second order random walk component ‘RW2MM‘ already contains trends for walking and cycling in combination with purpose ”Other”, and the added first order random walks are not well identified, giving rise to a slowly mixing MCMC simulation. The model predictions do show somewhat more volatile trends for cycling and walking. A disadvantage is that the OViN break for walking becomes smaller because of a higher 2009 estimate for the total number of walking trip legs pppd, which is not deemed very plausible given the expected higher amount of walking trips in 2010 due to the extremely high number of snow days in that year.

2. The same as the previous model extension where now the ‘RW2MM‘ component is restricted to have a common variance component for the (slope) innovations. This helps the MCMC simulation converge faster, but the resulting model fit is worse as indicated by higher values of the information criteria. The same implausible overall trend for walking is observed, as illustrated by Figure 6.5 (right panel).
3. Add first order random walks for walking and cycling for purpose ”Other”, but now replacing the corresponding second order random walks in the ‘RW2MM‘ component. Information criteria in this case are similar, but the same problem with less plausible OViN breaks still arises.
4. Instead of adding trend components we add random slopes for weather variables varying over the cross-classification of ageclass, purpose and mode. The weather variables considered are snow days (already used in the fixed effects part of the model), sun hours, fog days and precipitation days. All of these variables concern annual totals for the central measurement station in De Bilt. Each weather variable has been centered and standardized and given its own variance component. Correlations among the different weather random slope coefficients are allowed as well. Results show no improvement in information criteria, and hardly any observable changes in trend series. The estimated weather effects are very small, even though a few of them are significant with a plausible sign. This also doesn’t reduce the large estimated variance components for cycling and walking, purpose ”Other” in any significant way. This means that, even after removing weather effects, the resulting trend series for cycling and walking, purpose ”Other”, remains quite volatile.

In conclusion, even though some of these attempted model extensions may have slightly better information criteria, there is no compelling reason to adopt any of them. In the cases where a more appropriate flexible trend for walking and cycling is used the overall trend around 2010 for walking becomes less plausible. If, instead, weather variables are added as covariates in a random effects term, their effects are very small and hardly any change in trend plots can be discerned. In order to make progress on this issue it might be necessary to first deal with the implausibly high direct estimates for mode walking in 2010.

Regarding the 2010 high estimate for the number of walking trip legs, an analysis (Wüst, 2020) of the OViN 2010 survey weights showed that the first 7 days of the first three months of 2010 are quite strongly overrepresented, perhaps because of start-up issues of the OViN mixed mode design. These periods at least partly overlap with periods of snowy weather. Under such conditions it is known that people tend to favour walking, if possible, over some other modes, so it seems likely that the survey-weighted input figures for number of walking trips are over-estimated to some extent. To investigate this, a simple re-weighting has been carried out by which the 2010 OViN weights have been rescaled to match the proportions of the first seven days and the remaining number of days in each month. This re-weighting indeed yields a smaller value for the

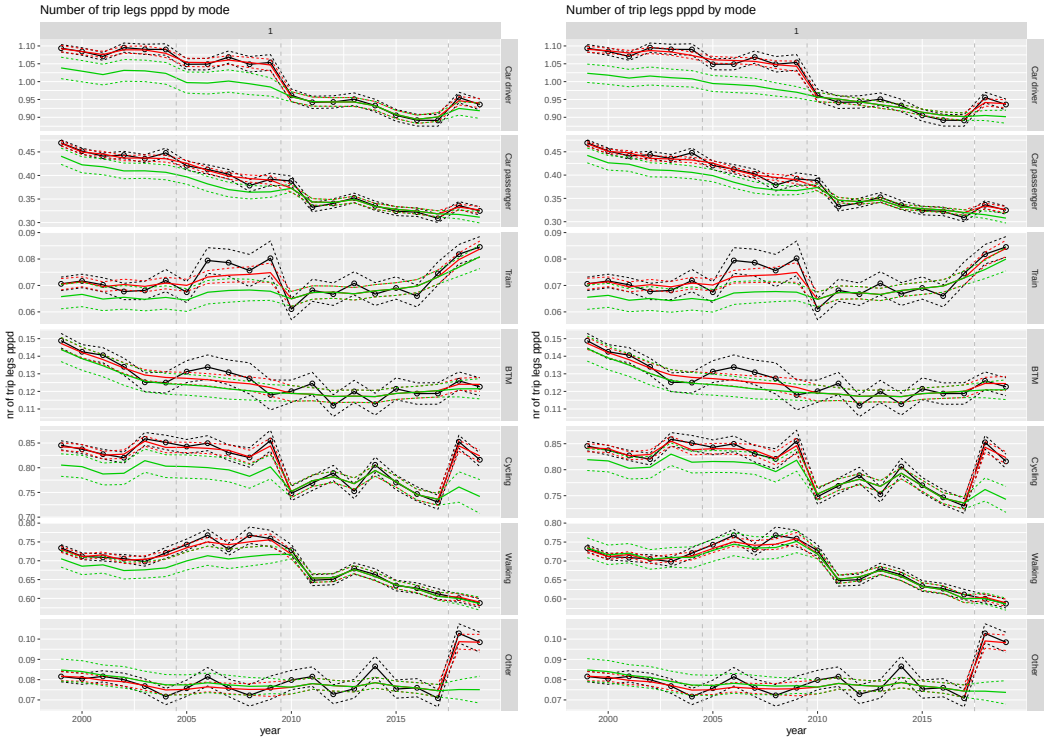

Figure 6.5 Comparison of total trip leg pppd estimates by mode between the selected model (left) and the same model with an added first order random walk for cycling and walking for purpose "Other", and a common variance parameter for the 'RW2MM' component (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

estimated total number of 2010 walking trip legs, although the difference is relatively small. Indeed it is so small that the trend estimates for walking for 2010 (and before) obtained by replacing the 2010 input estimates by the re-weighted estimates and refitting the model are only just visibly smaller, see Figure 6.6.

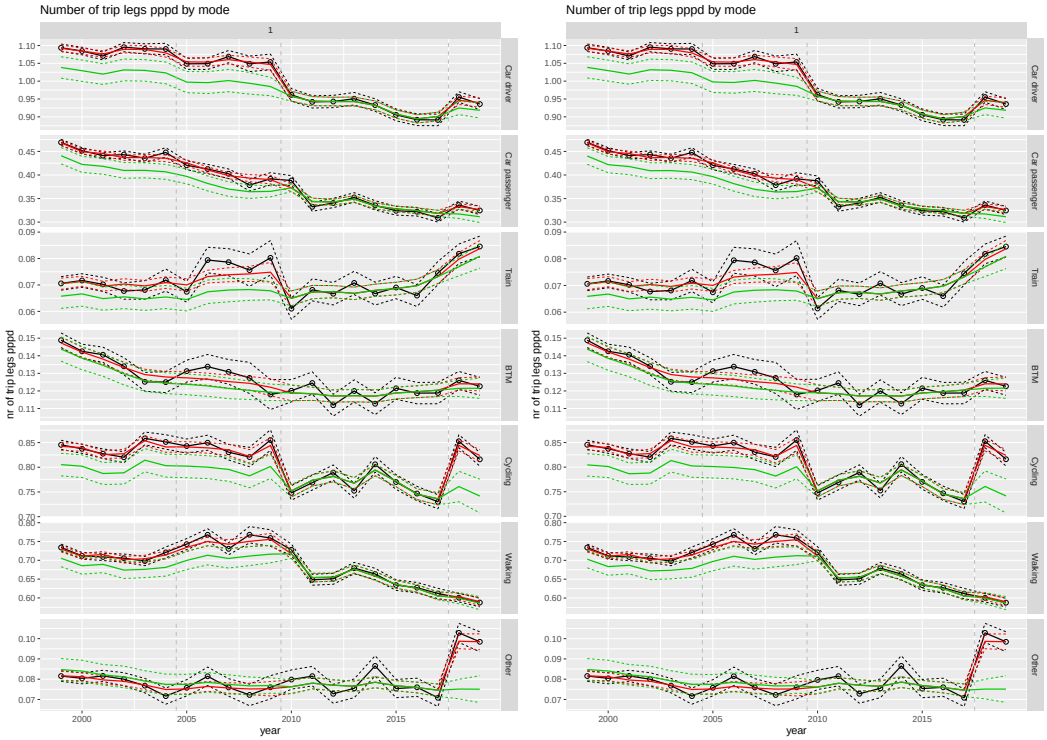

Figure 6.6 Comparison of total trip leg pppd estimates by mode for the selected model based on the original data (left) and on the data where 2010 is replaced by estimates based on a re-weighting by week within month (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

We have also tried a model in which all 2010 walking input estimates are treated as outliers, in a similar way as the 2009 estimates, cf. model component 'V_2009' in Table 5.1. However, the resulting changes are even smaller than in the case of the re-weighting just described. Also, a model has been fit to data excluding the 2010 initial walking estimates for purpose "Other", i.e. the latter have been treated as missing. This brings about small changes of size similar to those obtained by the re-weighting. Finally, a similar model in which all 2010 walking estimates are treated as missing was considered. This does result in a much smaller 2010 walking estimate, see (the right panel of) Figure 6.7. It also affects the overall OViN break for walking, which becomes larger, so that the trend estimates before 2010 become a little smaller.

Another potential sub-optimality in the current trend model is that the ordering of the age classes in the age classification variable is not used by the models described so far. It is possible to exploit the ordering, like the ordering in the years, by replacing some of the independent ageclass-dependent random effects in the model by correlated random effects. This way, the dependence of, for example, break coefficients can be made more smooth as a function of age(class), effectively borrowing strength from neighbouring age classes.

To investigate this, we tried several models. The best model thus obtained is a

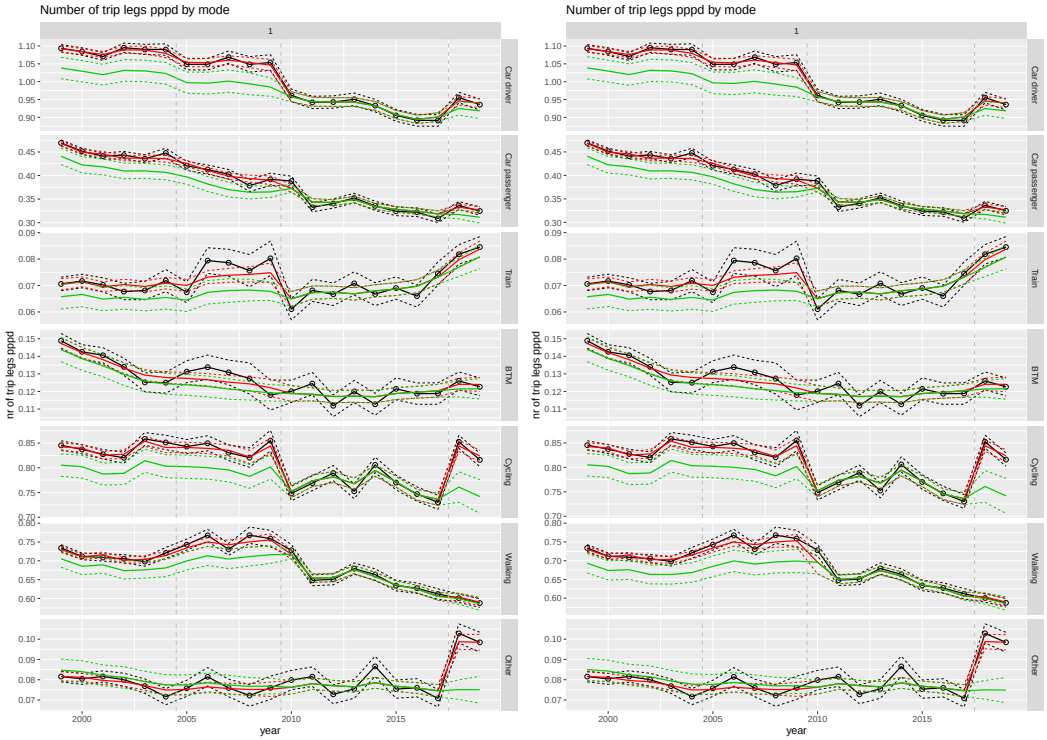

Figure 6.7 Comparison of total trip leg pppd estimates by mode for the selected model based on the original data (left) and on the data excluding all 2010 initial estimates for mode walking (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

modification of the selected model in which the independence of coefficients for different ageclasses in the 'V_BR' component is replaced by an autoregressive correlation structure over the ordered age classes. The autoregressive correlation parameter of this AR1 structure is set to 0.75 (having tried several values), meaning reasonably strong positive dependence between neighbouring age classes. The factor column in Table 5.1 for component 'V_BR' for this model is $sex * AR1(ageclass, 0.75) * purpose * mode$. At the same time the fixed effects for ODiN break effects have been extended in this model to include all such main effects for purpose and mode, i.e. the last line of 17 is replaced by $(purpose + mode) : br\_odin$. This resulted in slightly more plausible ODiN break estimates.

This model shows at least some improvements. Its model information criteria DIC and WAIC are more than 100 units below that of the model described in Subsection 5.1, indicating a better trade-off between model fit and model complexity. There are several easily visible differences in estimated trend series compared to the selected model, even at the highest aggregation levels, as displayed in Figures 6.8 - 6.10. One notices that the AR1 structure over age classes results in quite some differences in break estimates. In particular, it yields larger OViN breaks at the overall level, for purposes "Work", "Education" and "Other", as well as for most modes. Also the ODiN break estimates are somewhat larger. Note that the break estimates are not only larger but also more uncertain, thereby increasing the uncertainty about the (OViN-level) trend estimates outside the OViN period.

Figure 6.8 Comparison of total trip leg pppd estimates between the selected model (left) and the model with age class treated as ordinal (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure 6.9 Comparison of total trip leg pppd estimates by purpose between the selected model (left) and the model with age class treated as ordinal (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

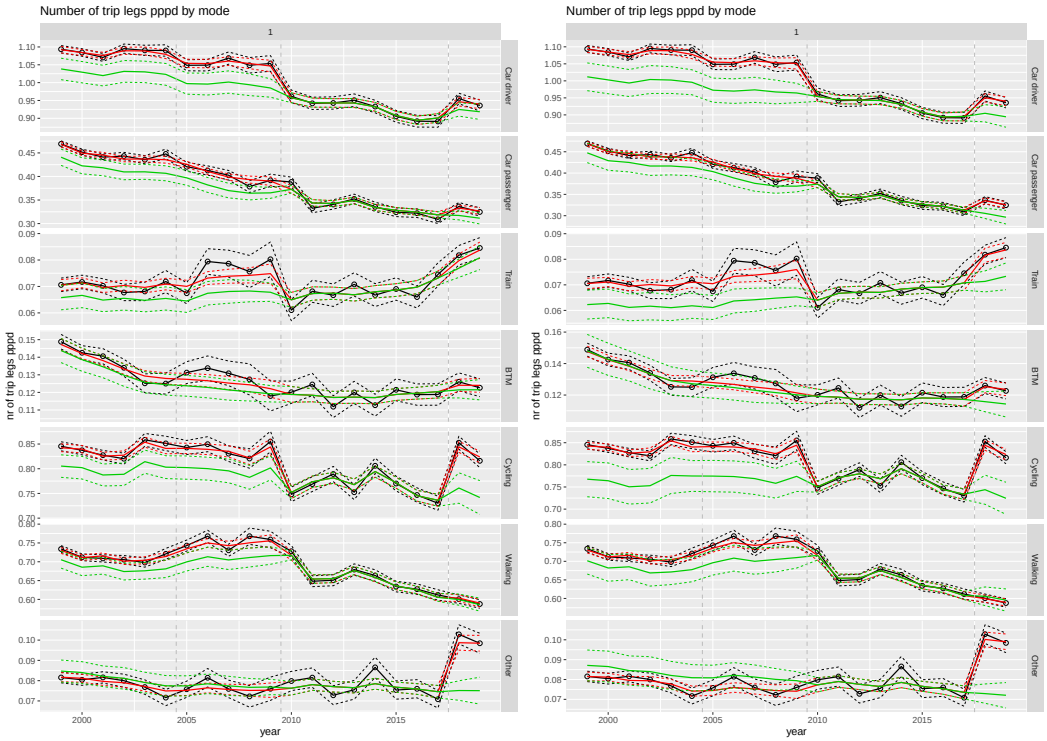

Figure 6.10 Comparison of total trip leg pppd estimates by mode between the selected model (left) and the model with age class treated as ordinal (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

6.3 Distance

Plots of trend estimates of mean distance per trip leg and mean distance per person per day are given in Appendices A.2 and A.3, respectively.

As with the number of trip legs, the importance of the additional global ODiN level break fixed effect included in (18) can be seen by comparing to the results based on the model without this additional effect, see Figures 6.11-6.13 for the highest aggregation levels. For aggregation the results for number of trip legs based on the model without the additional fixed effect ODiN breaks have been used in this case. From the figures it appears that without an overall ODiN break fixed effect some ODiN breaks are underestimated, resulting in large and implausible trend jumps.

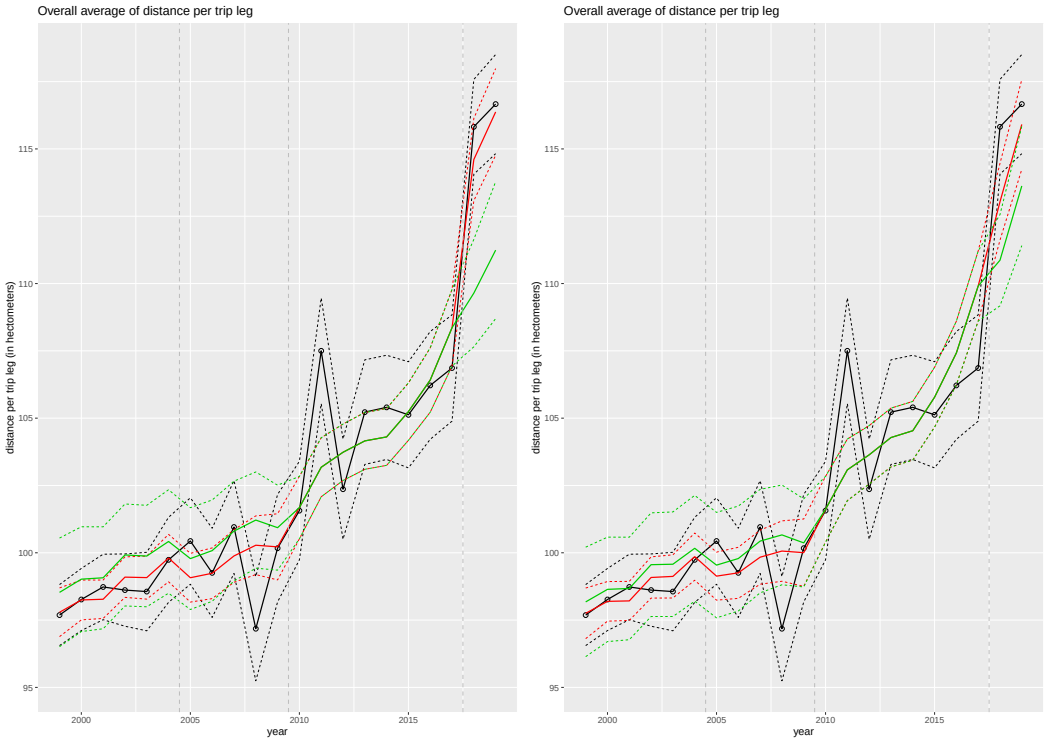

Figure 6.11 Comparison of total distance per trip leg estimates between the selected model (left) and the same model with fixed ODiN break effects removed (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

The direct estimates of distance per trip leg are rather volatile, even at the most aggregated level, see Figure A.70. The distance variable is also more affected by outliers, which occur in all years, and usually for domains with few observed trips. Therefore a Student-t distribution is used to fit the (log-transformed) distance variable. The posterior mean of the degrees of freedom parameter of the t distribution is 3.7 with a standard error of about 0.1.

Due to the noisier data, it is harder to detect fine changes in the underlying distance trends. In order to avoid overfitting, the model for distance is more parsimonious than that for the number of trip legs. One exception is that the distance model includes a fixed OViN break effect for mode ”Walking”. This effect was required to capture the very

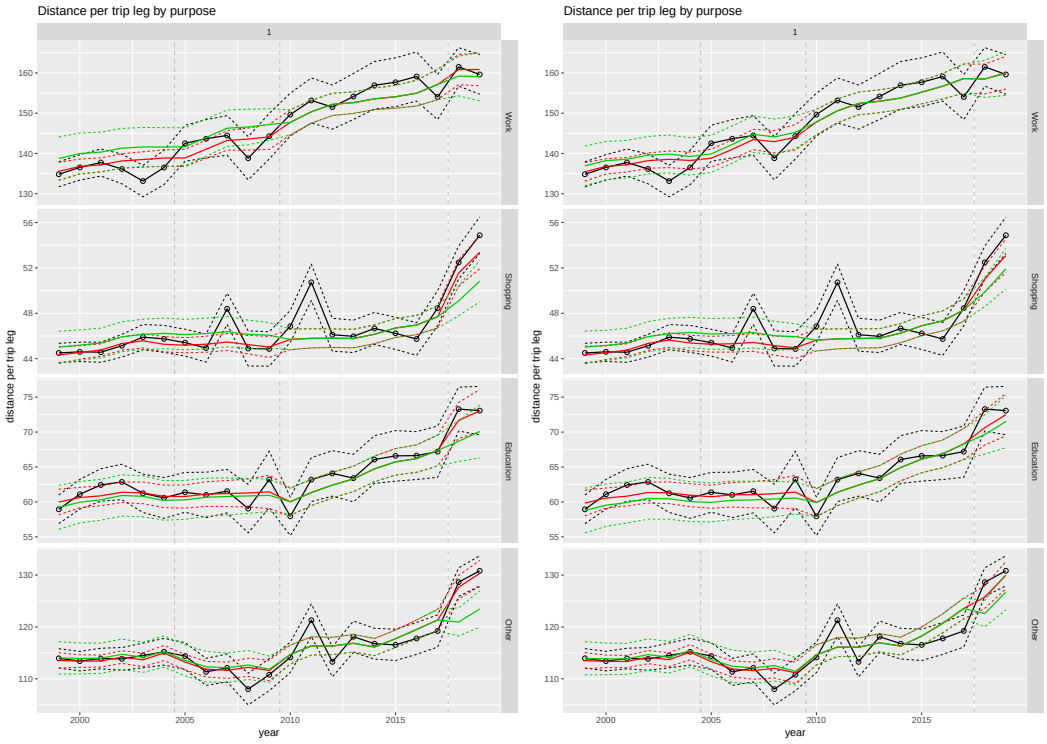

Figure 6.12 Comparison of total distance per trip leg estimates by purpose between the selected model (left) and the same model with fixed ODiN break effects removed (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

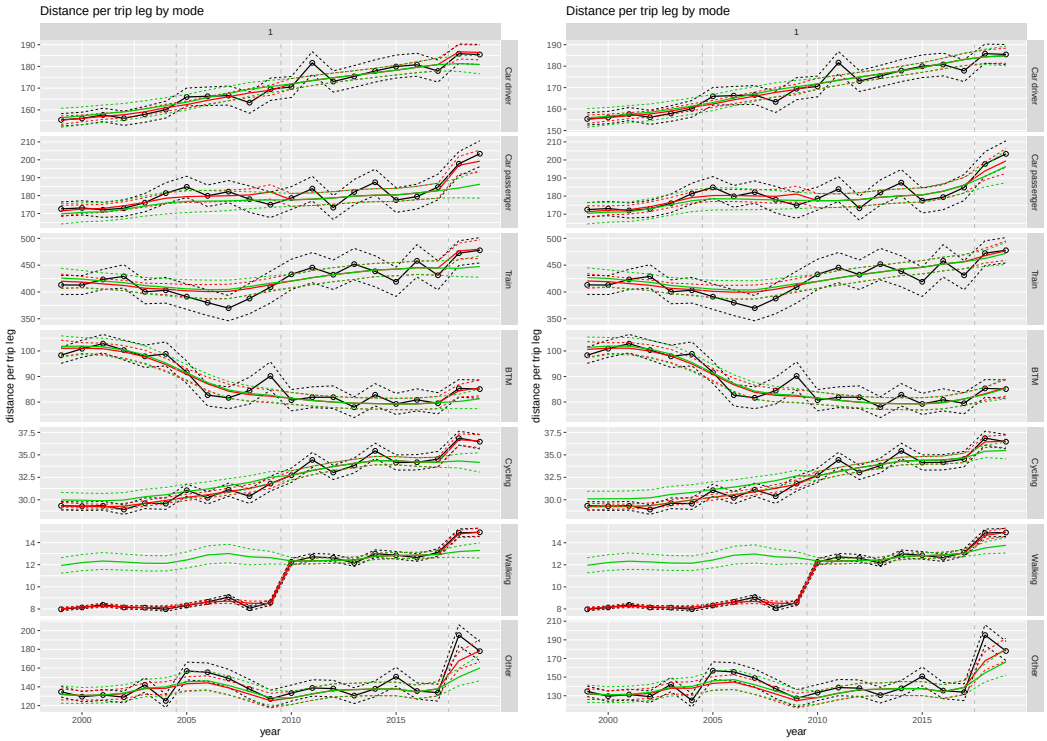

Figure 6.13 Comparison of total distance per trip leg estimates by mode between the selected model (left) and the same model with fixed ODiN break effects removed (right). Shown are direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

pronounced discontinuity in 2010 for mode ”Walking”[3)], as shown in Figure A.72.

Tables 6.3 and 6.4 list some parameter estimates (posterior means and standard errors). The ’V_BR’ component containing varying coefficients by domain for intercept, linear slope over time, and MON, OViN and ODiN breaks, shows, as in the trip legs model, negative correlation among the intercepts and OViN and ODiN breaks, see Table 6.3.

Table 6.4 shows that the mode-dependent scales of the smooth trend components by mode and purpose, as represented by model component ’RW2M’, are quite diverse. Here, the components are most volatile for ”BTM”, ”Walking” and, especially, ”Other”. As in the case of the trip leg model, the white noise component makes only a relatively small contribution to the trends.

	Intercept	yr.c	br_mon	br_ovin	br_odin
Intercept	25.2 (1.3)	6.1 (13.1)	5.1 (33.7)	-27.7 (9.8)	-27.4 (11.3)
yr.c		4.2 (0.7)	-3.0 (33.1)	8.9 (19.7)	17.7 (18.9)
br_mon			1.0 (0.8)	6.5 (33.7)	-11.3 (33.9)
br_ovin				11.6 (1.4)	3.0 (16.9)
br_odin					9.9 (1.3)

Table 6.3 Estimated standard deviations and correlations ($\times 100$) for the ’V_BR’ component.

Car Driver	Car Passenger	Train	BTM	Cycling	Walking	Other
1.9	5.0	10.1	19.5	8.1	26.3	63.9
(1.4)	(2.9)	(4.6)	(6.1)	(2.6)	(10.7)	(16.2)

Table 6.4 Estimated standard deviations ($\times 1000$) with their standard errors in parentheses under the diagonal covariance matrix for the ’RW2M’ component.

For distance too we have tried alternative models which exploit the ordering of age classes. In particular, an extension completely analogous to the one used for number of trips (using $sex * AR1(ageclass, 0.75) * purpose * mode$ in the component ‘V_BR‘ and $(purpose + mode) : br\_odin$ fixed ODiN break effects) was found to give similar improvements to the information model criteria and qualitatively similar changes to the estimated discontinuities. The models using ordinal ageclass will be investigated further in a possible follow-up of the trend estimation project.

7 Model assessment

As mentioned before, model selection was largely based on the WAIC, DIC, and graphical comparisons of the model fits and trend predictions at various aggregation levels. Besides, we investigate the bias of the model predictions, the variance reduction of the model predictions compared to the initial direct estimates, and the size of revisions if new data become available.

[3)] The effect is most probably due to the fact that in OViN walks are more often classified as single tours instead of consisting of a go and return trip.

7.1 Bias and variance reduction

Two discrepancy measures are defined to evaluate and compare the time series multilevel models. The first measure is the Relative Bias (RB) which expresses the differences between model estimates and direct estimates, as percentage of the latter. For a given model, the RB_{it} for a domain is defined as

$$RB_{it} = \frac{(\hat{\theta}_{it} - \hat{Y}_{it})}{\hat{Y}_{it}} \times 100\% . \tag{24}$$

with $\hat{\theta}_{it}$ the model prediction and $\hat{Y}_{it}$ the direct estimate for domain i and year t. This benchmark measure shows for each domain i how much the model-based estimates deviate from the direct estimates. The discrepancies should not be too large as one may expect that the direct estimates on average are close to the true average number of trip legs or distances. The second discrepancy measure is the Relative Reduction of the Standard Errors (RRSE) and measures the percentages of reduction in standard error of the model-based compared to the direct estimates, i.e.,

$$RRSE_{it} = 100\% \times (se(\hat{Y}_{it}) - se(\hat{\theta}_{it}))/se(\hat{Y}_{it}) . \tag{25}$$

Both measures are evaluated at different aggregation levels. The distributions of the measures are presented in terms of the minimum value, 1st quartile, median, mean, 3rd quartile and maximum value. The measures are presented for the following aggregation levels:

- Yearly estimates at the highest aggregation level (21 estimates, where (24) and (25) are averaged over all domains)
- Yearly estimates for the mode categories (21 by 7 estimates, where (24) and (25) are averaged over all domains that belong to a mode category)
- Yearly estimates for the purpose categories (21 by 4 estimates)
- Yearly estimates for the cross-classification of mode and purpose (21 by 7 by 4 estimates)
- Yearly estimates for the cross-classification of mode, purpose, gender and age-class (21 by 7 by 4 by 2 by 11 estimates)

7.2 Effect of parameterization of level breaks

As advised in Boonstra et al. (2019) we investigated to what extent the parameterization of the level break effects influences the trend series. To examine this, a different parameterization in which OViN is taken as the base level was tested. In this parameterization the 'V_BR' model component (see Tables 5.1 and 5.2) has been modified to include, besides intercepts and linear slopes, indicators for OVG, MON and ODiN periods, whose effects are allowed to vary over the full set of domains. No explicit OViN indicators are used in this case, as it is chosen to be the base level.

For both number of trip legs and distance, the models with this alternative parameterization of the 'V_BR' components' level breaks turned out to yield the same results (up to very small differences that can be attributed to Monte Carlo noise) as under the original parameterization described in Section 5. Intuitively, we believe that by using a full covariance matrix between the intercept, slope and different break effects, as we do in model component 'V_BR', the specific linear combinations of level breaks used do not matter.

7.3 Revision analysis

The models for the number of trip legs and distances are fitted to time series of different lengths, starting with series observed up until 2010 and then adding a year sequentially upto 2017. This is a so-called in real time analysis for the period 2010 to 2017. Since the revision analysis starts directly after the change-over from MON to OViN, this gives an impression how many observations under the new design are required until stable estimates for the discontinuities are obtained. For the last three years of the observed series under the OViN period, the discontinuity estimates have converged to a stable value, as will follow from the results in Subsections 7.4 and 7.5. For this period, this revision analysis gives an impression how stable the prediction for the last year is and how large its revision is when a new observation becomes available.

To evaluate the actual size of the revisions during the OViN period, absolute revisions (AR) and relative revisions (RR) for four different time lags or revision horizons ($k = 1, 2, 3, 4$ years) are calculated as $AR_k = |\hat{\theta}_{t|t+k} - \hat{\theta}_{t|t}|$ and $RR_k = |\hat{\theta}_{t|t+k} - \hat{\theta}_{t|t}|/\hat{\theta}_{t|t}$ respectively, where $\hat{\theta}_{t|t+k}$ denotes the estimate for θ_t based on a series observed until period $t + k$. The values of AR_k and RR_k are examined at the overall level and for the motive and mode levels separately. The values of AR_k and RR_k are are also calculated for the 28 levels of motive × mode levels. For convenience, the average values over the 28 levels are presented.

7.4 Trip legs

The distributions of the RB_{it} (24) and $RRSE_{it}$ (25) for the different aggregation levels specified in Section 7.1, are provided in Tables 7.1 and 7.2, respectively. The bias at the highest aggregation level is negligible (average of 0.97%) and on average has only a slightly larger size at the more detailed levels. The reduction of the variance is the smallest at the highest aggregation level (22.16% on average) and gradually increases with the level of detail to 48.42% at the most detailed level.

Level	Min.	1st.Qu.	Median	Mean	3rd.Qu.	Max.
Overall	-0.33	0.43	0.91	0.97	1.38	3.31
Motive	-5.84	-2.11	0.91	0.77	3.79	7.12
Mode	-13.64	0.10	4.19	1.37	5.69	7.44
Motive and Mode	-29.36	-0.51	4.85	1.63	6.64	7.53
Sex, Ageclass, Motive, and mode	-62.87	-1.31	-0.19	-1.20	0.31	65.21

Table 7.1 Summary statistics of relative bias (in %) at different annual aggregation levels for the SAE estimates of mean number of trip legs pppd

Level	Min.	1st.Qu.	Median	Mean	3rd.Qu.	Max.
Overall	11.38	18.49	23.65	22.16	26.09	33.93
Motive	18.73	24.21	26.98	28.31	31.07	40.54
Mode	19.13	20.97	33.03	28.65	34.59	37.25
Motive and Mode	9.40	28.59	33.33	32.35	36.31	53.58
Sex, Ageclass, Motive, and mode	32.17	43.72	47.46	48.42	50.90	91.38

Table 7.2 Summary statistics of relative reduction of standard errors (in %) at different annual aggregation levels for the standard errors of the SAE estimates of mean number of trip legs pppd

The size of the revisions is evaluated graphically at the highest aggregation level in Figure 7.1. The left panel of Figure 7.1 illustrates the revisions if directly after the change-over

from MON to OVIN trends are published at the OVIN level for time series observed up until 2010 and then adding one year at a time. The right panel of Figure 7.1 shows the revisions if trends are published at the OVG level in real time for the considered period 2010 to 2017. The sizes of the revisions are also evaluated at the purpose and mode levels in Figures A.208 and A.209.

Figure 7.1 Signal of mean number of trip legs for overall level measured at OVIN and OVG levels

In real time, estimates for the discontinuity due to the change-over from MON to OViN are presented in Figure 7.2. This figure illustrates that in the first 4 years the revisions of the discontinuity estimates are quite large. Also after four years, the estimates are still subject to small revisions. The updates for the OViN discontinuities during the years directly after the implementation of OViN are one factor that causes the revisions of the trends visible in 7.1. The estimates of the OVIN discontinuity are also calculated for the purpose and mode levels and are shown in Figures A.210 and A.211. Figure A.210 shows that the discontinuity estimates stabilize after four years for work and shopping purposes, while the estimates steadily decrease to zero for the education purpose and steadily increase for the purpose "Other". At the mode level, the discontinuity estimates shown in Figure A.211 tend to stabilize after 3-4 years of OViN design for all transportation mode except for the mode "Walking".

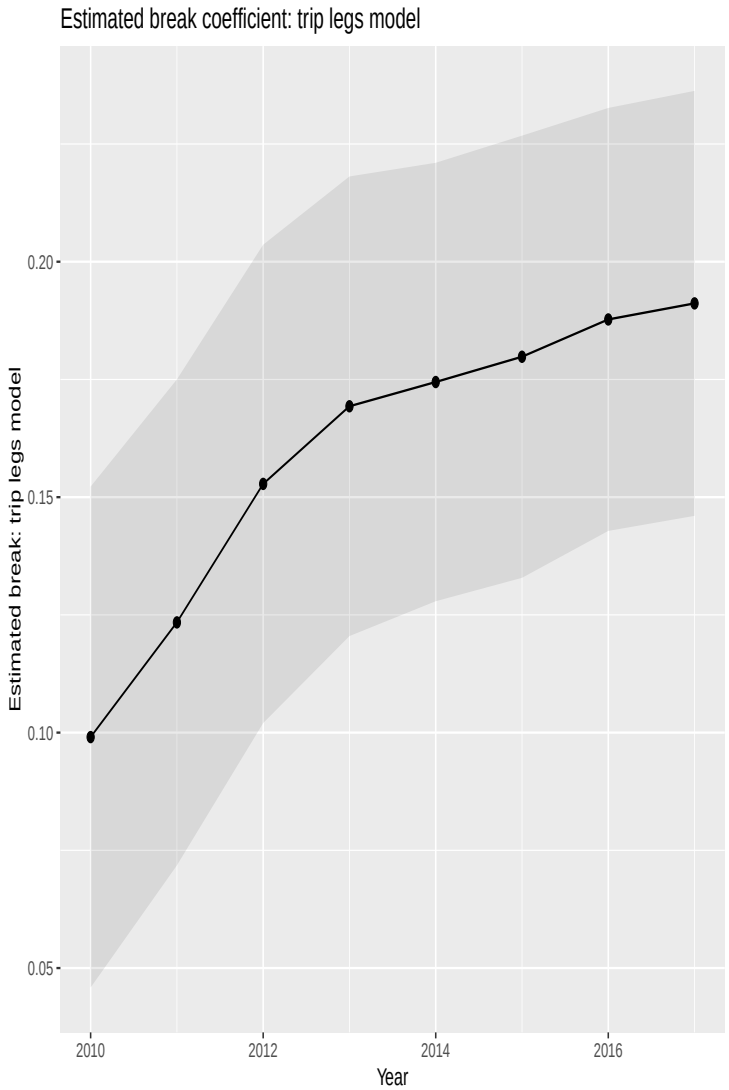

Figure 7.2 In real time estimates from 2010 to 2017 for the discontinuity for number of trip legs at the overall level with approximate 95% confidence interval

Recall that the level intervention variable for the MON breaks equal one during the MON period 2004-2009 and equals zero elsewhere. The level intervention variable for the OViN break equals one starting from 2010, and zero before. If signals are estimated at

the OVG level, no discontinuities are added to the signal at all time periods. If signals are estimated at the OViN level, the estimate for the OViN discontinuity is added to the signal at all time periods. The left panel of Figure 7.1 shows large revisions during the OVG and MON period and relatively small revisions during the OViN period, if the signals are estimated at the OViN level. The right panel of Figure 7.1, where signals are estimated at the OVG level, shows hardly any revision during the OVG period, small revisions during the MON period and larger revisons during the OViN period. During the MON period, the estimate for the MON discontinuity is removed. The relatively small revisions of the signals over the MON period (2004-2009) if the signals are estimated at the OVG level, are the revisions of the MON discontinuity estimates if additional observations under OViN become available. This comparison suggests that revisions of the signals due to unstable discontinuity estimates directly after the change-over to a new survey design result in large revisions in the periods where the actual survey design used to collect data differs from the level of the survey design for which the signals are estimated. Or to say it differently, the level of the signal estimates is most accurately identified for the period where the survey design used for the data collection coincides with survey design at which the signal estimates are produced.

These findings are also confirmed by the in real time analyses at the level of the purpose and mode categories in Figures A.208 and A.209 respectively. The Figures also illustrate that the size of the revisions vary over purpose and mode level, as for example the revisions are comparatively lower for purpose ”Work” and for mode ”Train” and ”Other”.

To compare the size of the revisions during the OViN period when signals are estimated at OViN level with the size of the revisions if the signals are estimated at OVG level, the values of AR_k and RR_k are calculated for four different revision horizons ($k = 1, 2, 3, 4$ years) in real time for the period after the OViN design starts. Figure 7.3 (top panel) shows the values of AR_k for the overall level signals of the number of trip legs pppd published at OViN and OVG level during the OViN period. Table A.1 also shows AR_k and RR_k for the overall level signals of the number of trip legs pppd. It seems that the values of RR_k provide very similar findings as those of AR_k values. The average values of AR_ks calculated for the signals of purpose×mode levels are also shown in the bottom panel of Figure 7.3.

The comparison of AR_k as well as RR_k at the overall level in Table A.1 for various revision horizons $k = 1, 2, 3, 4$ indicates that neither OViN nor OVG level estimates provide smaller revisions during the OViN period. However, the bottom panel of Figure 7.3 shows that the size of the revisions, averaged over motive and mode level values, are smaller for all k and t when the trends are published at OViN level. This indicates that the size of revisions at the disaggregated levels are smaller when the publication level is OViN. The purpose and mode level signals shown in Figure A.208 and Figure A.209 respectively indicate that the size of the revisions at OViN level are comparatively smaller than those at the OVG level during the OViN period. The values of AR_k for the purpose and mode level signals shown respectively in Figure A.212 and Figure A.213 confirm these findings. Figure A.212 shows that the AR_k values at the OVIN level are smaller in most of the cases of k and t for all the motives compared to those AR_k values at the OVG level. While Figure A.213 shows similar scenarios for most of the mode categories except for the mode BTM.

So the purpose, mode, and purpose×mode level signals indicate that estimates at the OViN level are subject to smaller revisions during the OViN period compared to estimates at the OVG level. As mentioned before, it takes about 4 years until stable estimates for

the discontinuities are obtained (Figure 7.2). In the period after 2014 the revisions are very small. This is an indication that once stable estimates for the discontinuities are obtained, the revisions of the model predictions are small. This might suggest that, once sufficient data are available after the last redesign, no revision policy is required.

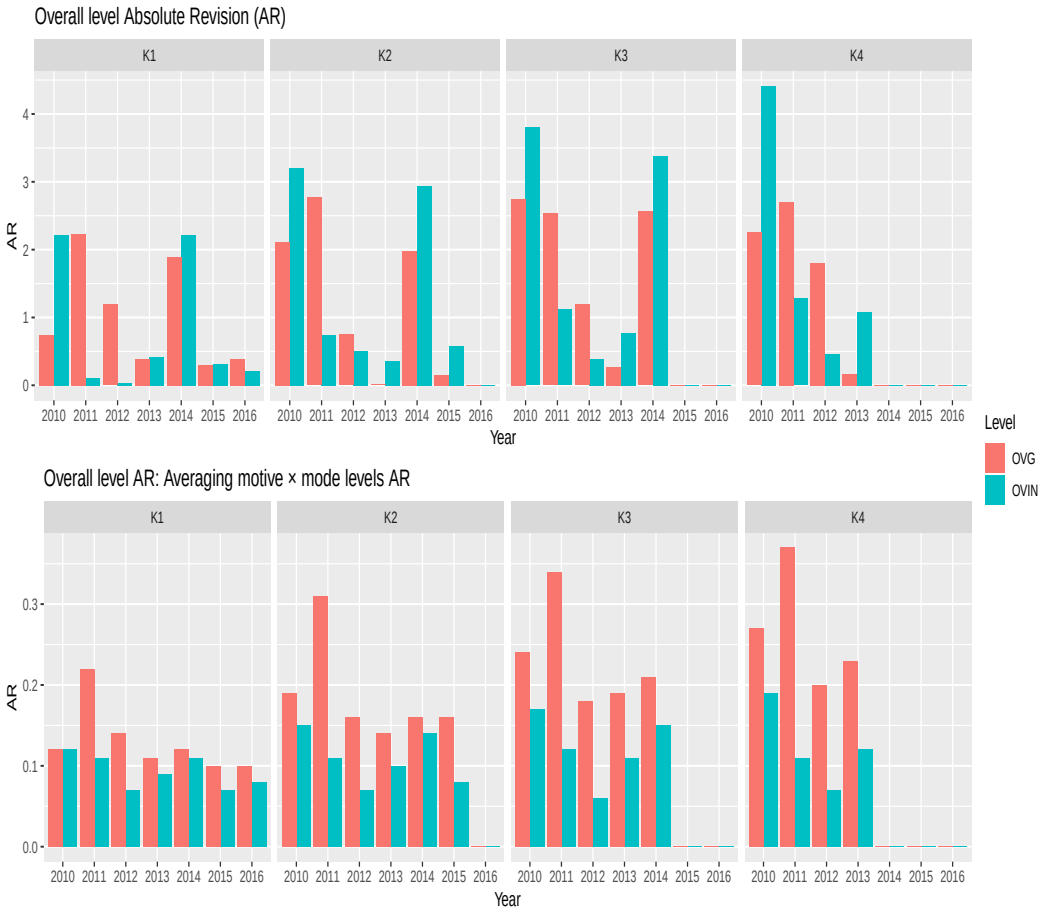

Figure 7.3 Absolute Revision ($AR_k \times 100$) of the signals at overall level (top panel) and the average values of ARs calculated for the motive × model level signals (bottom panel) for mean number of trip legs pppd when the signals are published at OViN and OVG level for revision horizon $k = 1, 2, 3, 4$.

7.5 Distance

The distributions of the RB_{it} (24) and $RRSE_{it}$ (25) for the different aggregation levels specified in Section 7.1, are provided in Tables 7.3 and 7.4, respectively. The bias at the highest aggregation level is negligible (average of 2.23%) and gradually increases to an average of -3.53% at the most detailed level. The reduction of the variance is the smallest at the highest aggregation level (35.26% on average) and gradually increases with the level of detail to 58.15% at the most detailed level.

The size of the revisions for distance per trip leg is also evaluated graphically first at the highest aggregation level in Figures 7.4. The left panel of Figure 7.4 shows the revisions if trends are published at the OViN level in real time analysis for the period 2010 to 2017, while the right panel of Figure 7.4 illustrates the revisions if trends are published at the OVG level. The size of the revisions for the purpose and the mode levels are plotted in Figure A.214 and Figure A.215 respectively considering trends are published at the OViN and OVG levels.

Level	Min.	1st.Qu.	Median	Mean	3rd.Qu.	Max.
Overall	-1.05	1.67	2.31	2.23	2.94	5.46
Motive	-0.64	-0.30	1.11	2.19	3.60	7.18
Mode	-1.51	-0.24	0.78	1.23	1.42	6.99
Motive and Mode	-15.20	-1.08	-0.14	0.15	0.96	16.46
Sex, Ageclass, Motive, and mode	-57.94	-4.82	-0.66	-3.53	0.35	51.71

Table 7.3 Summary statistics of relative bias (in %) at different annual aggregation levels for the SAE estimates of mean distance per trip legs pppd

Level	Min.	1st.Qu.	Median	Mean	3rd.Qu.	Max.
Overall	12.52	33.19	40.48	35.26	41.06	43.03
Motive	30.44	33.51	35.21	36.44	38.14	44.92
Mode	7.62	24.13	36.53	33.24	44.37	51.54
Motive and Mode	7.44	26.64	42.27	38.76	52.90	58.31
Sex, Ageclass, Motive, and mode	-95.28	55.54	60.08	58.15	64.41	73.21

Table 7.4 Summary statistics of relative reduction of standard errors (in %) at different annual aggregation levels for the standard errors of the SAE estimates of mean distance per trip legs pppd

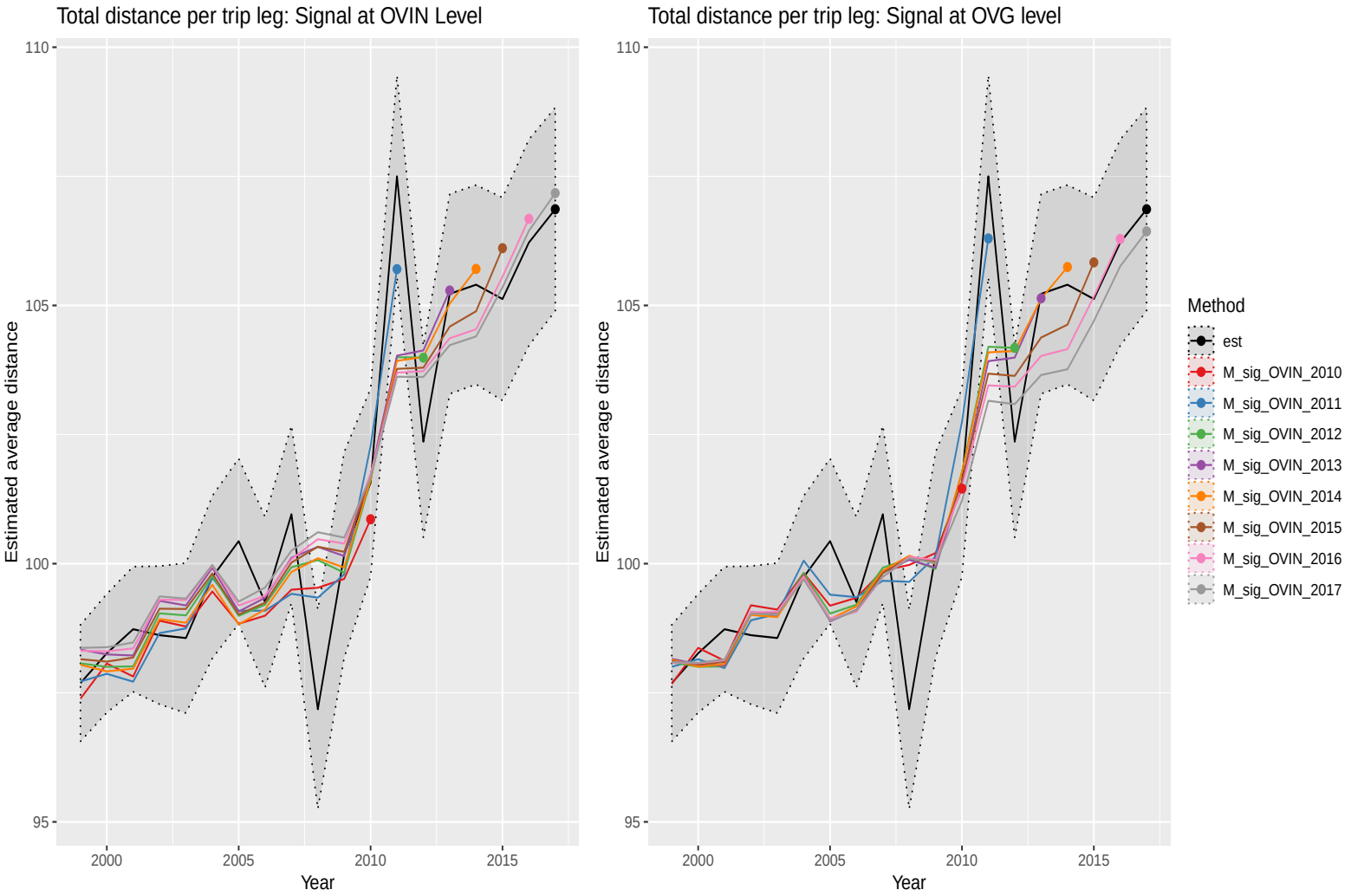

Figure 7.4 Signal of average distance of trip legs for overall level measured at OVIN and OVG levels

In real time, estimates for the discontinuity due to the change-over from MON to OViN are presented in Figure 7.5. This figure illustrates that the discontinuity for distance per trip leg at the aggregated level is not significantly different from zero. The estimates of the OVIN discontinuity are also calculated for the purpose and the mode level and are shown in Figures A.216 and A.217 respectively. Figure A.216 shows irregular patterns for the discontinuity estimates for all purposes except ”Work”. These discontinuity estimates, however, are not significantly different from zero. Discontinuity estimates for the mode categories are given in Figure A.217. It can be seen that discontinuity estimates decrease steadily four years after the OViN design and then stabilizes around a negative value for mode ”Car driver”, ”BTM”, ”Cycling”, and ”Walking”. As expected discontinuities are not significantly different from zero for all modes expect for ”Walking”, for which discontinuity estimates are found always lower than zero.

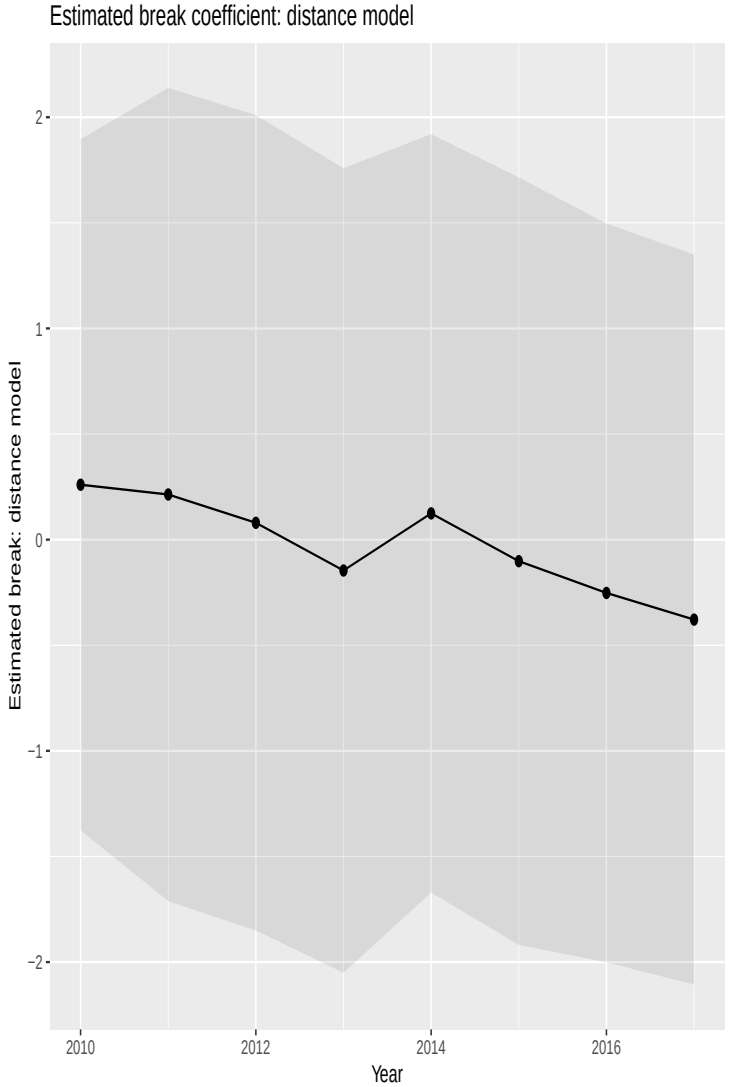

Figure 7.5 In real time estimates from 2010 to 2017 for the discontinuity for average distance of trip legs at the overall level with 95% confidence interval

The revisions of the signals visible in Figure 7.4 are most likely caused by the revisions of the predicted number of trip legs pppd, which are used in the aggregation of the average trip distances. As in the case of the number of trip legs, the revisions for the distance are larger for periods where the survey design for which predictions are made, deviates

from the actual survey design used to collect the data.

For the OViN period, it appears that the revisions at the aggregate level are slightly larger if the predictions are at the OVG level than those at the OViN level. The comparison of AR_k for various k and t shown in the top panel of Figure 7.6 indeed indicates that the OViN level revisions are consistently smaller for all revision horizons and years except for the year 2010.

The bottom panel of Figure 7.6 also shows that, at the level of the cross-classification of mode and motive, the revisions for the estimates at the OViN level are smaller for most of the combinations of k and t except for the year $t = 2010$. This Figure illustrates that the size of revisions at OViN level are smaller also at the disaggregated levels during the OViN period.

The values of AR_k are also calculated for the signals of mean distance per trip leg by the purpose and mode categories and are shown in Figure A.218 and A.218 respectively. For motive categories, the AR_k values for estimates at the OViN level are smaller than the values for the estimates at the OVG level for all the motives in most combinations of t and k except for purpose " Shopping". This finding is supported by Figure A.214.

At the mode level, the values of AR_k shown in Figure A.218 seem very similar whether the signals are published at either OVIN or OVG level during the OViN period with some exception. Such results are in line with Figure A.215. The estimated discontinuities which are not significantly different from zero may be one of the reasons for this indifference in the size of revisions for the distance model. Thus, the values of AR_k at motive and mode level indicate that the revisions of the estimates at OViN level at the overall level are only slightly smaller mainly due to smaller revisions at the motive level when signals are published at OViN level instead of OVG level.

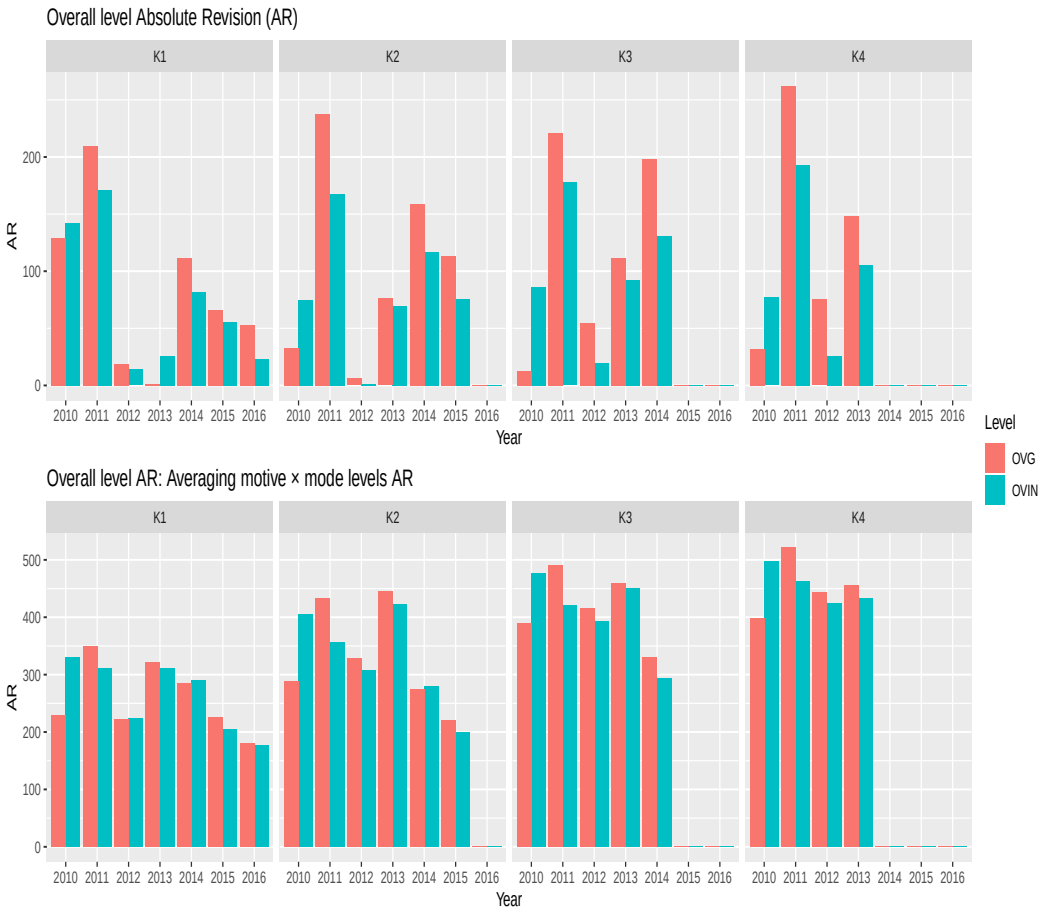

Figure 7.6 Absolute Revision ($AR_k \times 100$) of the signals at overall level (top panel) and the average values of AR_ks calculated for the motive × model level signals (bottom panel) for mean distance per trip leg when the signals are published at OViN and OVG level for revision horizon $k = 1, 2, 3, 4$.

8 Discussion

In Boonstra et al. (2019), two models have been developed for estimating trends for mobility indicators based on the Dutch Travel Survey (DTS) data observed from 1999 until 2017. In 2018, the survey process transited from OViN to ODiN. In this paper the models are extended to use additional data of 2018 and 2019, observed under the latest DTS survey design, i.e. ODiN. The target variables modeled are the average number of trip legs per person per day and the average distance per trip leg for domains that are defined by a cross-classification of age, gender, purpose, and transportation mode at a yearly frequency. In the process, the age-classification has been extended from 9 to 11 classes. Furthermore some potential model improvements are investigated in this paper.

In the first stage direct estimates are compiled from the DTS data as well as their standard errors, using the general regression estimator. The direct estimates are the input for the time series models and are first transformed to better meet normality assumptions. For the number of trip legs a square-root transformation is used and for distance a log transformation. The standard error estimates are also transformed and subsequently smoothed with a generalized variance function model.

The resulting direct estimates at the level of the aforementioned cross-classification are used as input for multilevel time series models, which are fitted using MCMC simulations. The models account for discontinuities due to three redesigns that occurred due to change-over from OVG to MON in 2004, the change-over from MON to OViN in 2010 and the change-over from OViN to ODiN in 2018. Discontinuities are predominantly modeled as random effects to avoid overestimation. The DTS time series are also affected by outliers. The model for trip legs contains random effects to absorb the most dominant outliers in 2009, while the model for distances assumes a Student-t distribution for the sampling errors. Other important model components are smooth temporal trends at different levels of the cross-classification variables. Some of the random effects are assigned non-normal priors to induce a stronger form of regularization. Furthermore, observations of the number of annual snow days are used as a covariate series in the trip legs model.

Model predictions and trends for the number of trip legs per person per day, distance per trip leg and distance per person per day are obtained at different aggregation levels of the cross-classification by aggregating the model predictions and trends at the most detailed level. Currently only two years of data under the new ODiN design are available, which means that the ODiN discontinuity estimates are not yet stable and subject to revisions if additional years become available. Therefore it was decided for now to base the trends at the OViN level. This means that the trends are derived from the model predictions by including the OViN break effects and excluding the MON and ODiN break effects.

Model selection is based on model information criteria (WAIC and DIC), and plausibility analyses using external information including subject matter knowledge provided by KiM and Rijkswaterstaat about mobility in the Netherlands. In Boonstra et al. (2019) the adequacy of the selected models was confirmed using a set of model diagnostics, including posterior predictive checks, showing no obvious signs of bias, overfitting or underfitting in the majority of the time series. The model-based estimates are roughly 25% to 50% more precise than the direct estimates in the case of trip legs and 35% to 60% in the case of distance, from the highest aggregation level to the most detailed level.

An in-real-time-analysis of the discontinuities has been conducted for the OViN period, i.e. from 2010 up until 2017, at high aggregation levels. It is found that stable estimates for the discontinuities are obtained after about 4 years in the case of trip legs, while for distance no significant discontinuities are observed. The analysis illustrates that the revisions of signals are large for periods where the survey design that is actually used to collect the data does not coincide with the level at which the signals are estimated. For periods where the survey design used to observe the series coincides with the level of the survey design at which the signals are estimated, the combination of the trend and the discontinuity is better identified compared to the other periods. As a result, the revisions of the signals for this period are the smallest. Once the discontinuity estimates are sufficiently stable, the revisions of the signal predictions will become small.

It appears that the ODiN breaks are significantly different from zero and have substantial magnitudes. The results observed under ODiN turn out to be more similar to the levels observed under OVG and MON. It was nevertheless decided to estimate, for the time being, the signals at the OViN level. This choice implies that revisions for the OViN period 2010-2017 can be expected to be small and that larger revisions can be expected for the ODiN period, once new data for 2020 and beyond become available.

A number of potential model improvements have been investigated in this report. First it is analyzed how to handle the unexpected large estimates for ”Walking” in 2010. Due in part to a large amount of snow days in the first quarter, the observed number of ”Walking” trip legs is large. As a result the OViN break appears to be underestimated. Adding auxiliary weather variables does not solve this issue. It is also investigated to what extent a different weighting scheme can solve this problem, but this only has a small effect on the direct estimates. Modeling ”Walking” in 2010 as an outlier with random effects hardly improves the results. A potential solution is to treat all observations related to ”Walking” as missing (which comes down to modeling them as outliers with fixed effects at the most detailed cross-classification). This does yield a clear improvement of the total trend estimates for ”Walking”.

For modes ”Walking” and ”Cycling” and for motive ”Other”, the observed time series are very volatile, which was a reason to consider alternative trend models for these classes. Eventually, this did not yield any substantial improvements to the model fits. Another direction of model development did give rise to some potential improvements: here the ordering of the age classes was used to obtain random effects for discontinuities that vary more gradually over subsequent age classes. In particular, a first order autoregressive (AR(1)) dependence was assumed. Together with additional fixed ODiN break effects for purpose and mode this led to improved model information criteria as well as many changes to the OViN and ODiN discontinuity estimates, at least some of which appear to be improvements. Whether these models, for both trip legs and distance, are overall an improvement remains to be seen. This will be investigated further in a possible extension of the trend estimation project.

It can be expected that the model for 2020 needs further adaptation since the lockdown due to the corona crisis will have a substantial effect on the evolution of the time series. Another point for further research is to investigate alternative model specifications for the discontinuities. In particular, the AR(1) specification for discontinuities between age classes and treating ”Walking” observations in 2010 as missing require further model evaluation and plausibility analyses. Finally, a further revision analysis of the ODiN discontinuity estimates has to be conducted to facilitate the decision when the level of signal estimates can be changed from the OViN level to the ODiN level.

References

Boonstra, H. J. (2020). *mcmcsae: Markov Chain Monte Carlo Small Area Estimation*. R package version 0.5.0.

Boonstra, H. J. and J. van den Brakel (2018). Hierarchical bayesian time series multilevel models for consistent small area estimates at different frequencies and regional levels. Statistics Netherlands discussion paper, December 4, 2018.

Boonstra, H. J., J. van den Brakel, and S. Das (2018). Computing input estimates for time series modeling of mobility trends. CBS report.

Boonstra, H. J., J. van den Brakel, and S. Das (2019). Multi-level time series modeling of mobility trends - final report. CBS Discussion Paper, 30 October 2019.

Brezger, A., L. Fahrmeir, and A. Hennerfeind (2007). Adaptive gaussian markov random fields with applications in human brain mapping. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 56*(3), 327–345.

Carter, C. K. and R. Kohn (1996). Markov chain monte carlo in conditionally gaussian state space models. *Biometrika 83*(3), 589–601.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Datta, G. S. and P. Lahiri (1995). Robust hierarchical bayes estimation of small area characteristics in the presence of covariates and outliers. *Journal of Multivariate Analysis 54*(2), 310–328.

Fabrizi, E., M. R. Ferrante, and C. Trivisano (2018). Bayesian small area estimation for skewed business survey variables. *Journal of the Royal Statistical Society Series C 67*(4), 861–879.

Fabrizi, E. and C. Trivisano (2010). Robust linear mixed models for small area estimation. *Journal of Statistical Planning and Inference 140*(2), 433–443.

Fay, R. and R. Herriot (1979). Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association 74*(366), 269–277.

Gelfand, A. and A. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1*(3), 515–533.

Gelman, A. and J. Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*(4), 457–472.

Geman, S. and D. Geman (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell. 6*, 721–741.

Juárez, M. A. and M. F. J. Steel (2010). Model-based clustering of non-gaussian panel data based on skew-t distributions. *Journal of Business and Economic Statistics 28*(1), 52–66.

Konen, R. and H. Molnár (2007). Onderzoek verplaatsingsgedrag - methodologische beschrijving. Statistics Netherlands.

Lang, S., E.-M. Fronk, and L. Fahrmeir (2002). Function estimation with locally adaptive dynamic models. *Computational Statistics 17*(4), 479–500.

Molnár, H. (2007). Mobiliteitsonderzoek nederland - methodologische beschrijving. Statistics Netherlands.

O'Malley, A. and A. Zaslavsky (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association 103*(484), 1405–1418.

Park, T. and G. Casella (2008). The bayesian lasso. *Journal of the American Statistical Association 103*(482), 681–686.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rao, J. and I. Molina (2015). *Small Area Estimation*. Wiley-Interscience.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.

Särndal, C.-E., B. Swensson, and J. Wretman (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. *Biometrika 76*, 527–537.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. Springer.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B 64*(4), 583–639.

Tang, X., M. Ghosh, N. S. Ha, and J. Sedransk (2018). Modeling random effects using global–local shrinkage priors in small area estimation. *Journal of the American Statistical Association*, 1–14.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

van den Brakel, J., G. Griffiths, T. Surzhina, P. Wise, J. Blanchard, X. Zhang, and O. Honchar (2017). A framework for measuring the impact of transitions in official statistics. ABS research paper 1351.055.158, Australian Bureau of Statistics.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research 11*, 3571–3594.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research 14*, 867–897.

West, M. (1984). Outlier models and prior distributions in bayesian linear regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 431–439.

Willems, R. and J. van den Brakel (2015). Methodebreukcorrectie ovin. PPM 210514/12, Statistics Netherlands.

Wolter, K. (2007). *Introduction to Variance Estimation*. Springer.

Wüst, H. (2020). 2010 lopen. short note (in Dutch).

Appendix

A Time-series plots model-based and direct estimates

A.1 Number of trip legs per person per day

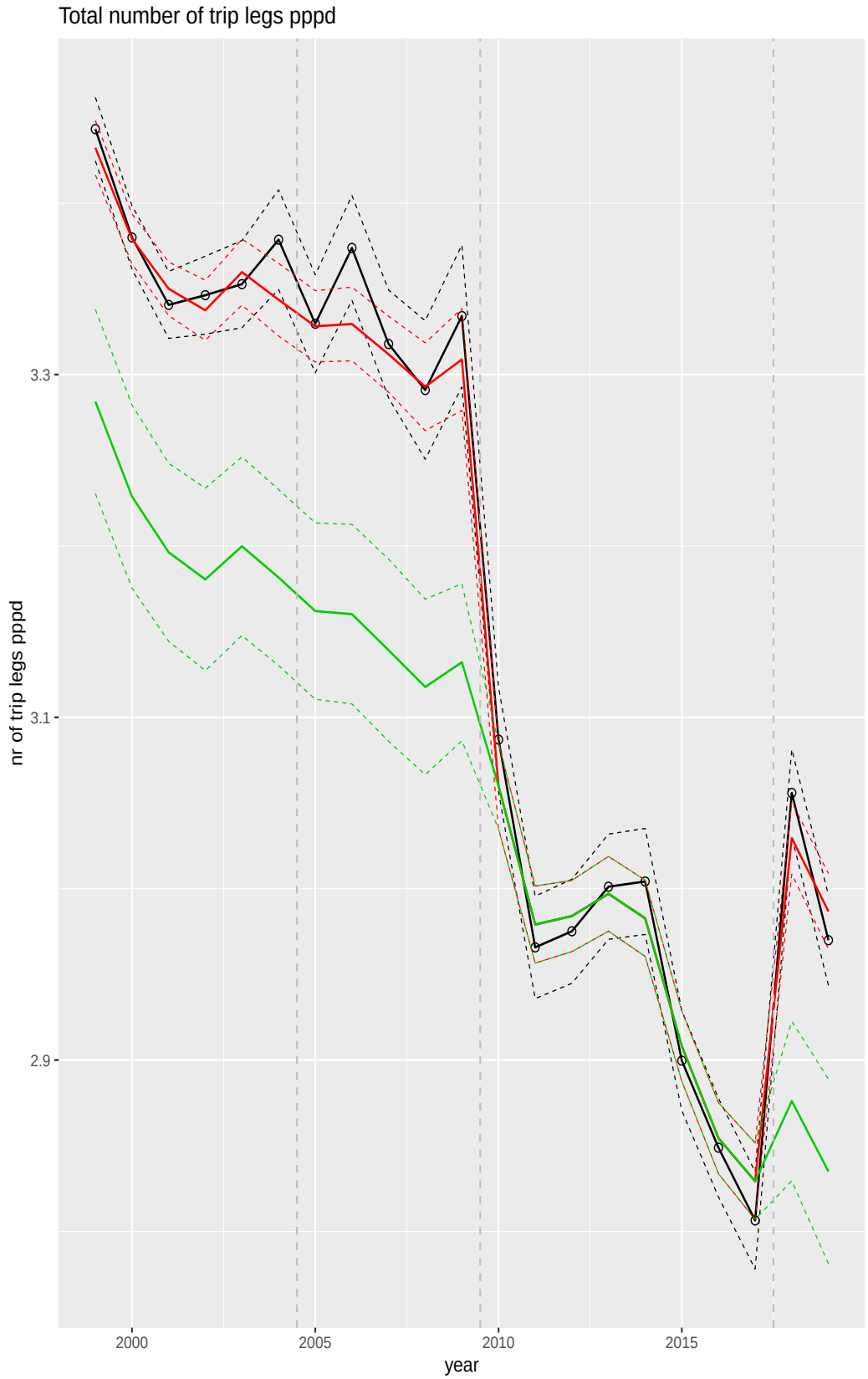

Figure A.1 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

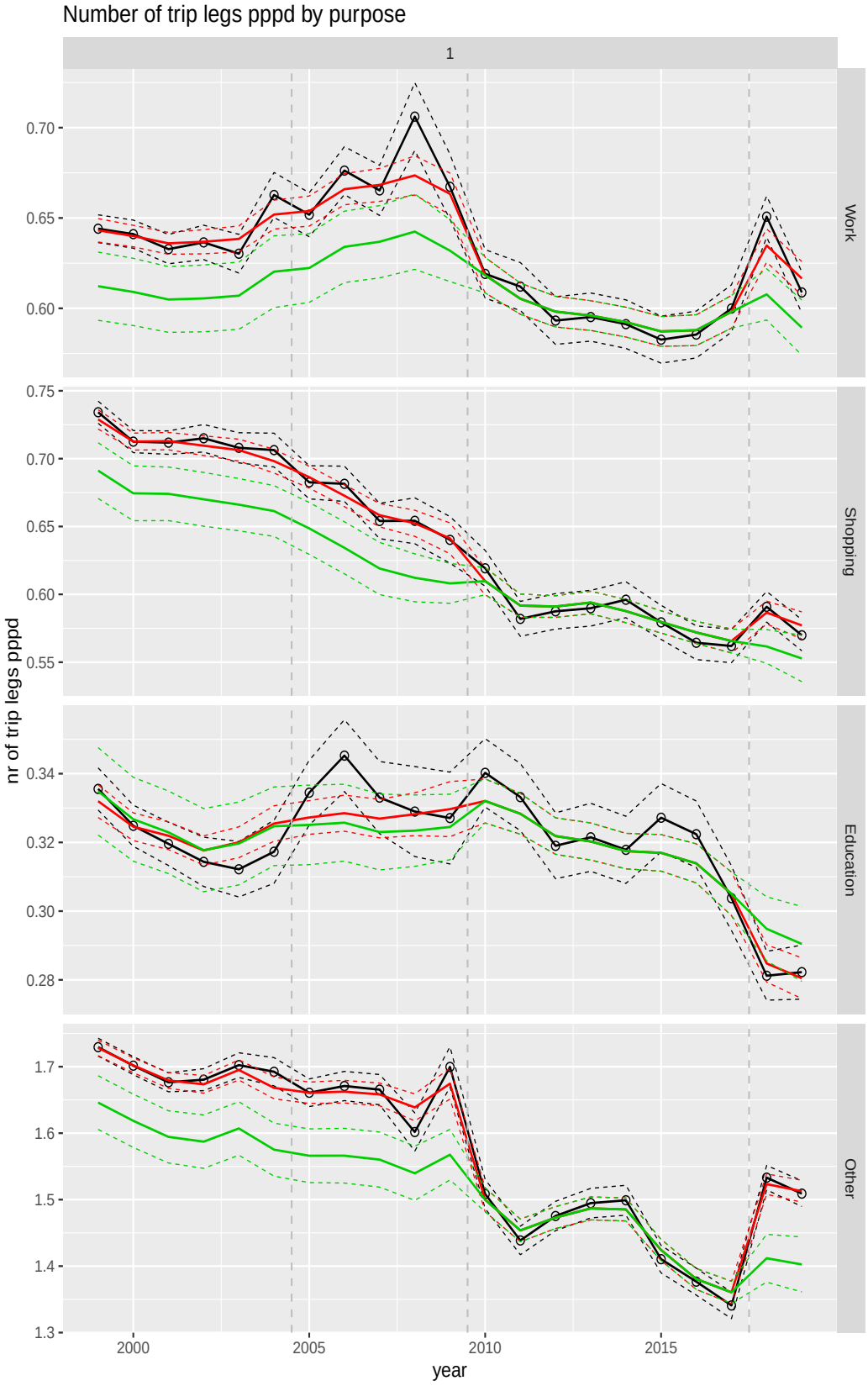

Figure A.2 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

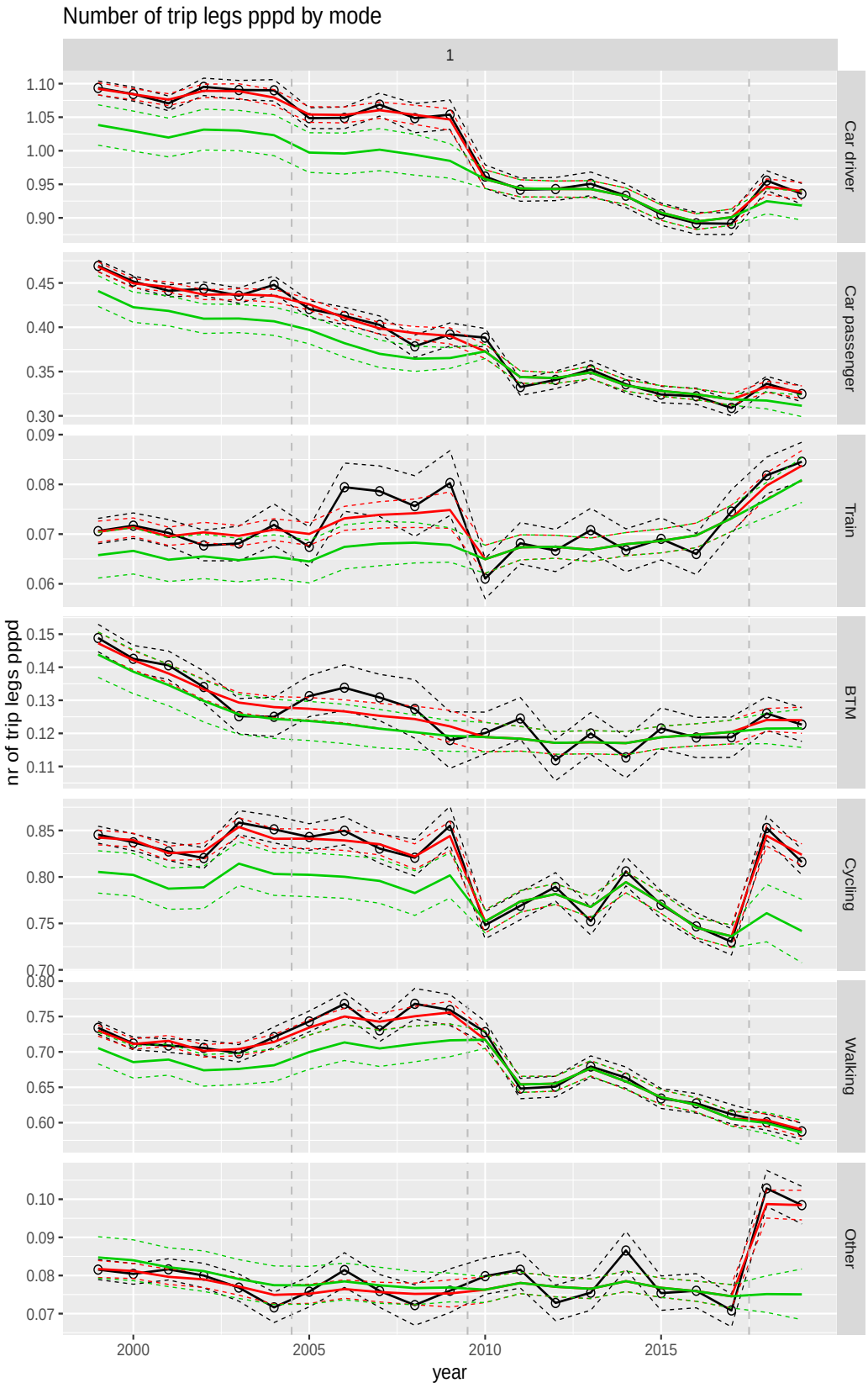

Figure A.3 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

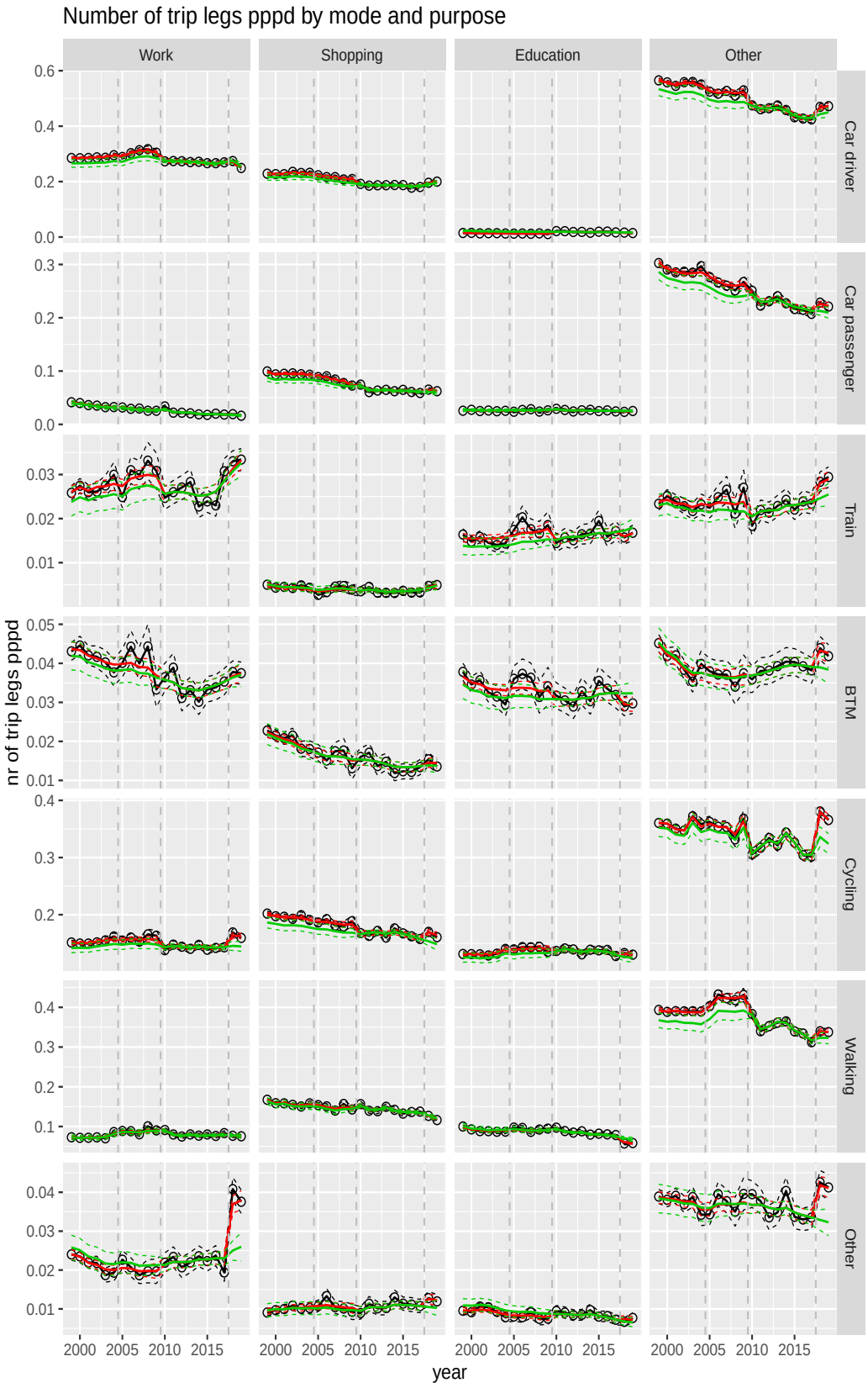

Figure A.4 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

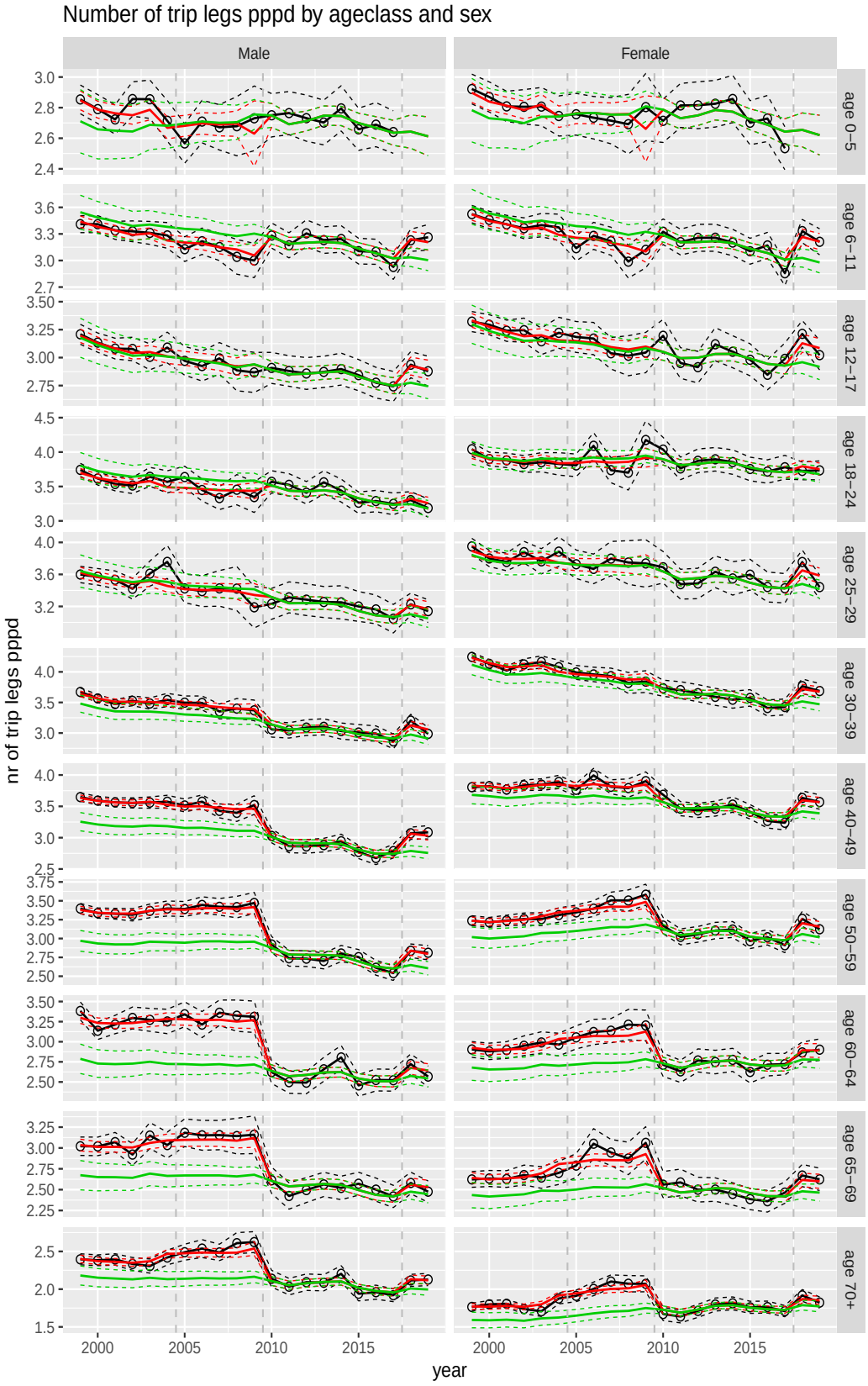

Figure A.5 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

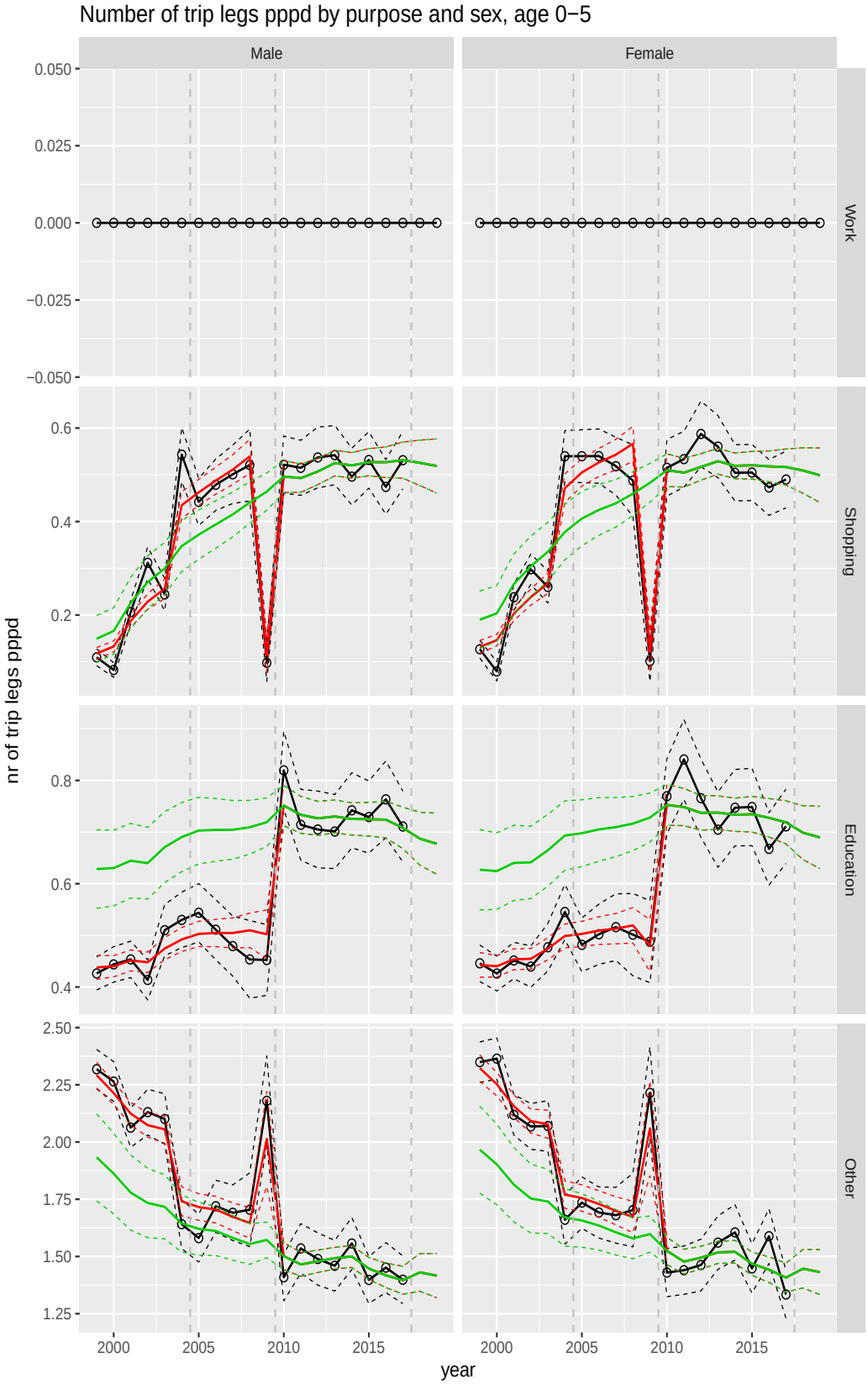

Figure A.6 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure A.7 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

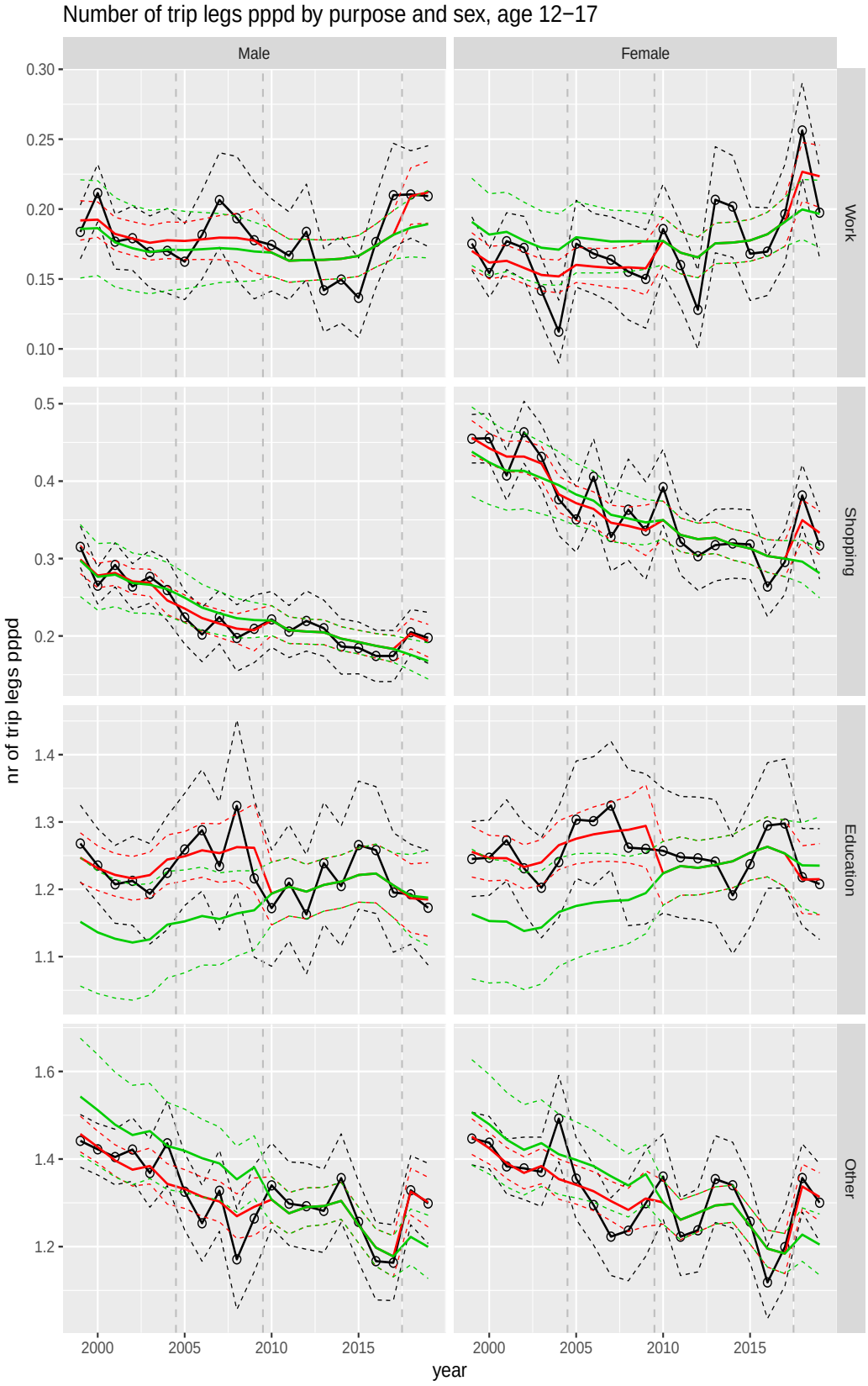

Figure A.8 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

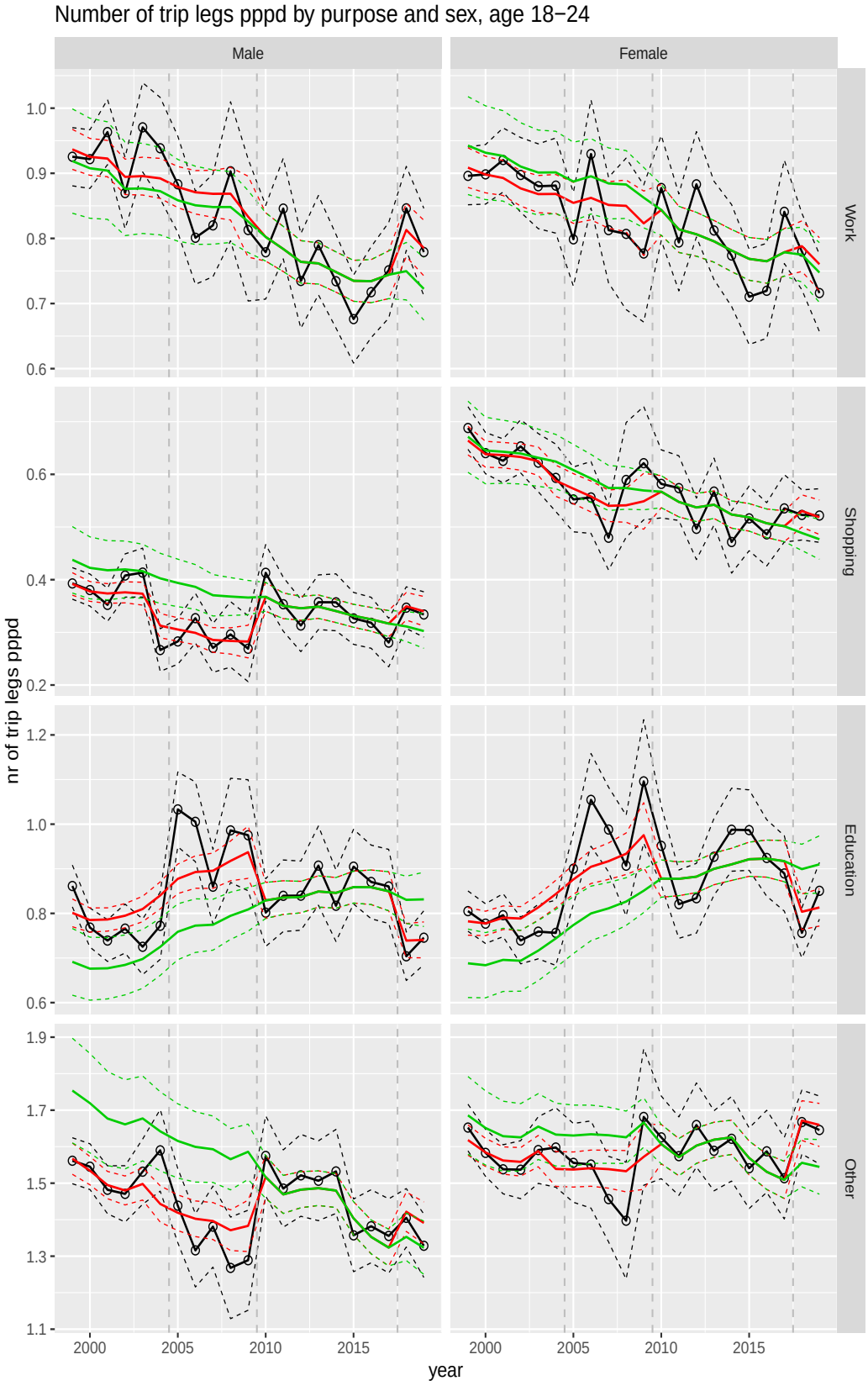

Figure A.9 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

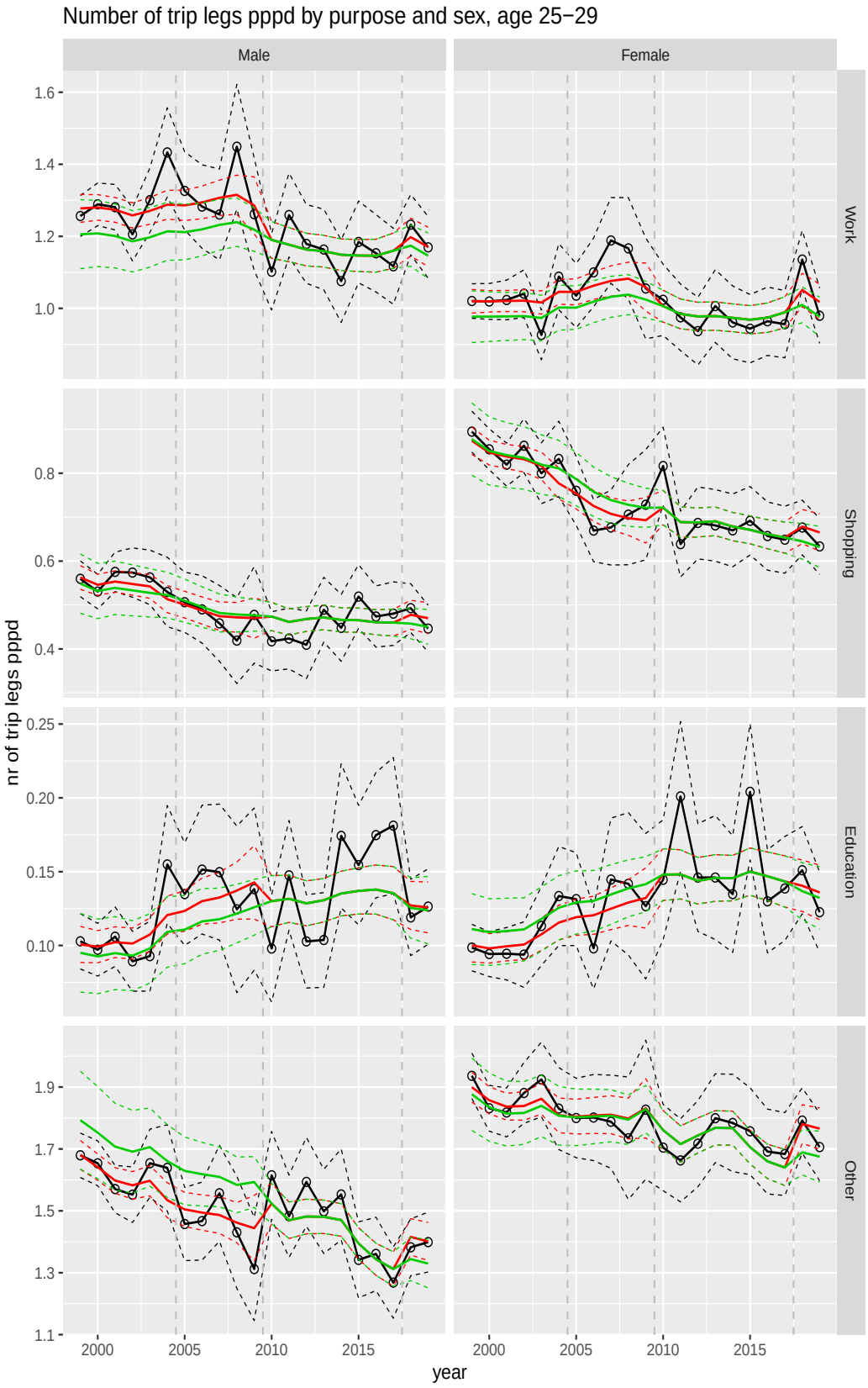

Figure A.10 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

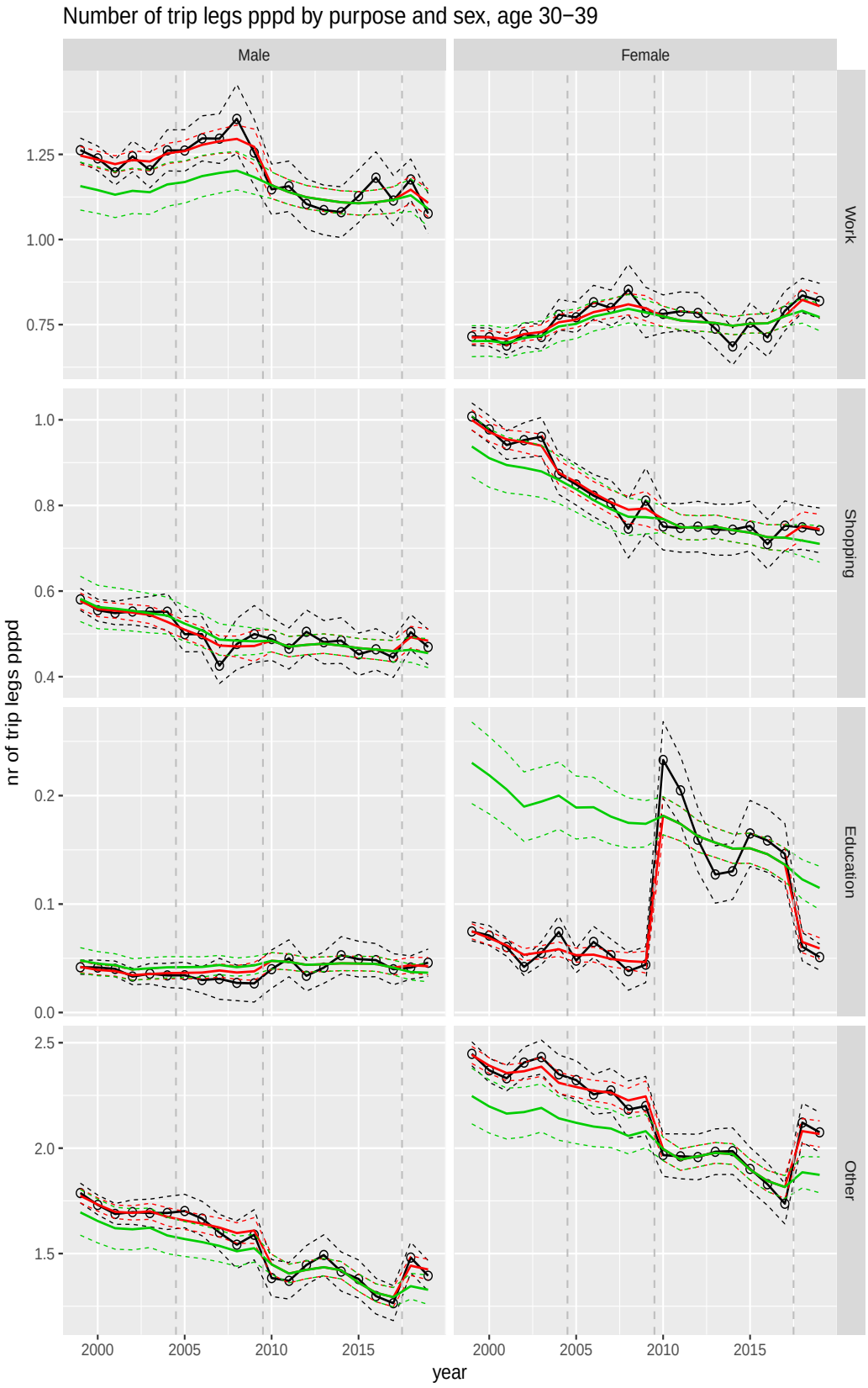

Figure A.11 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

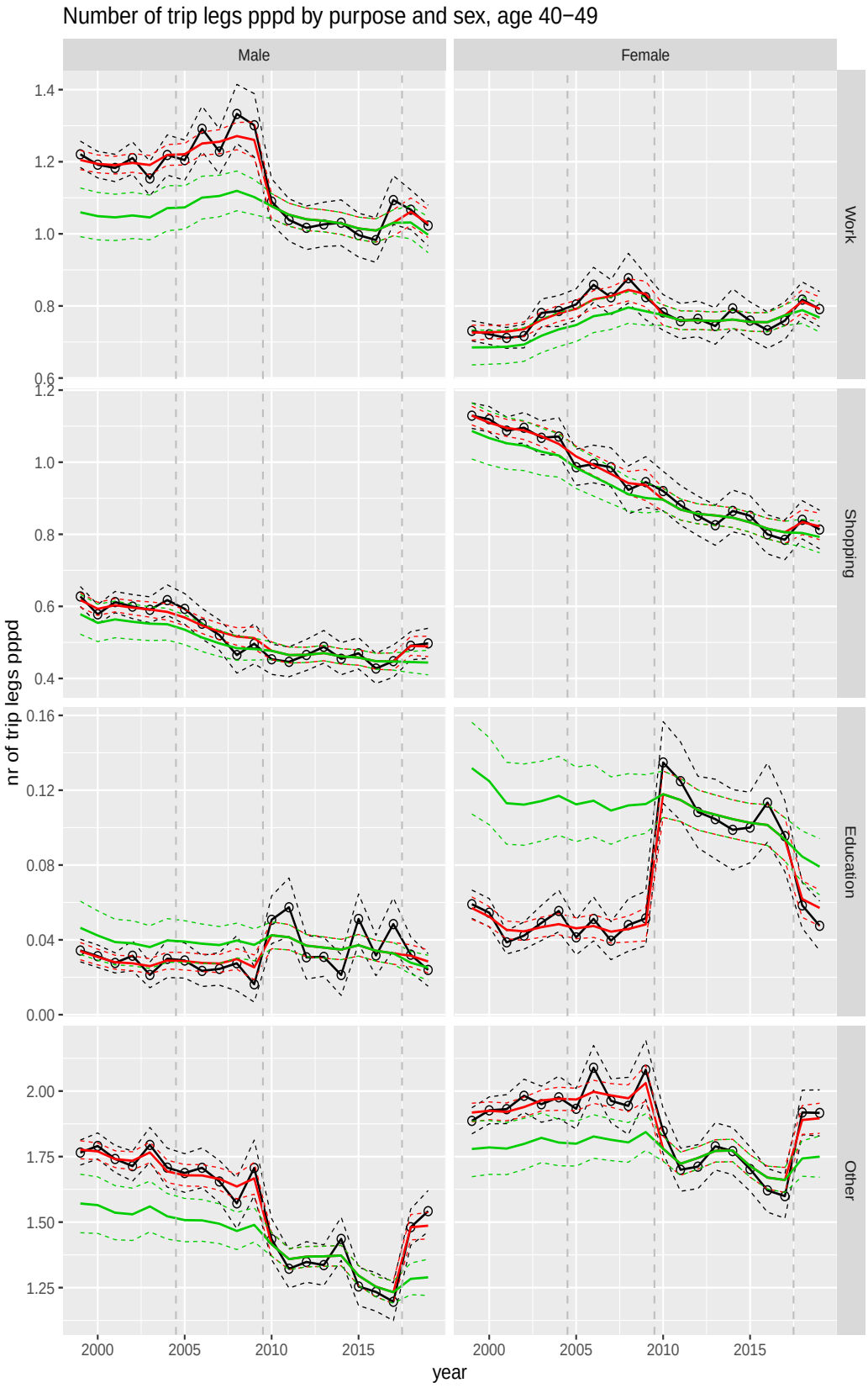

Figure A.12 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

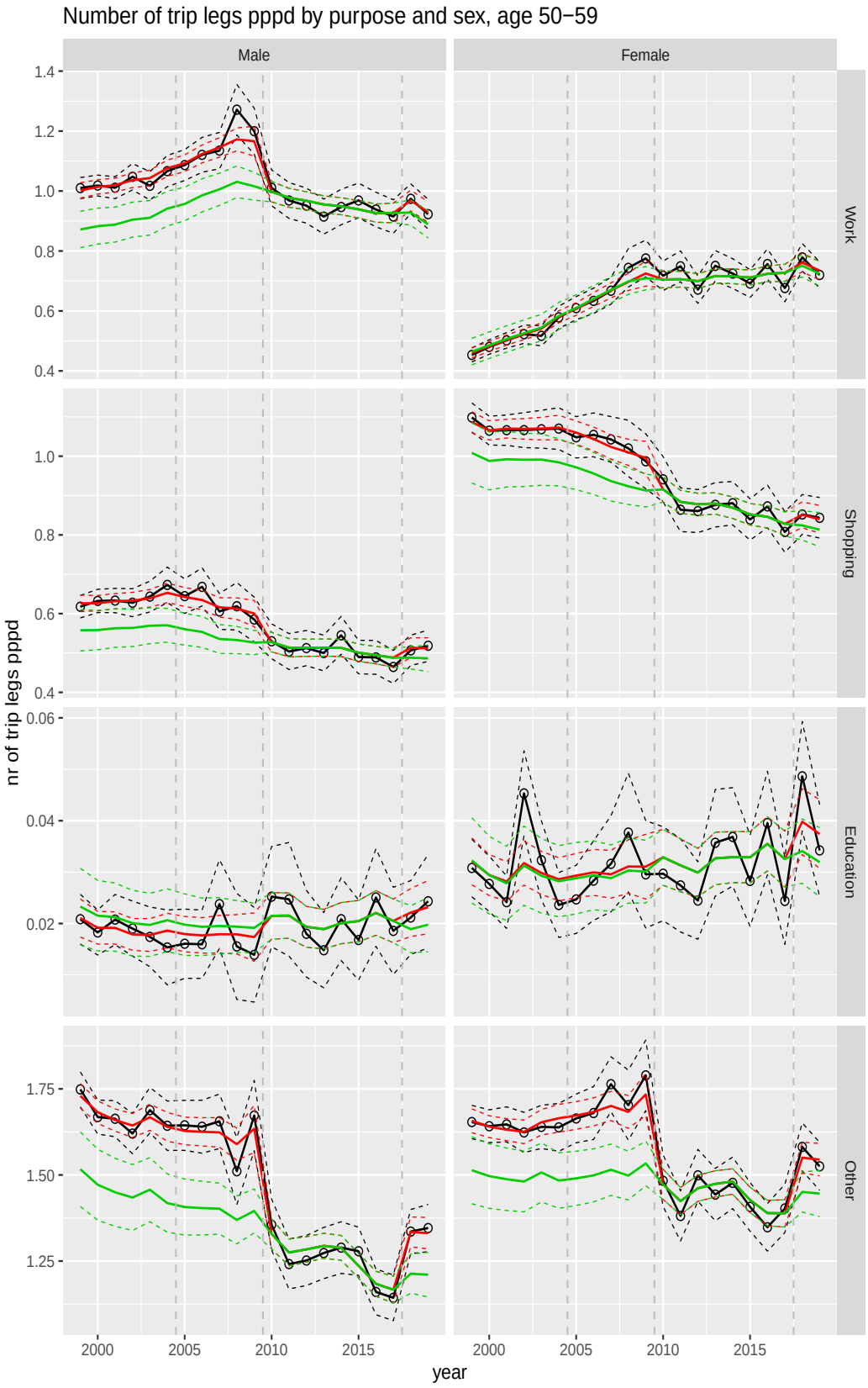

Figure A.13 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

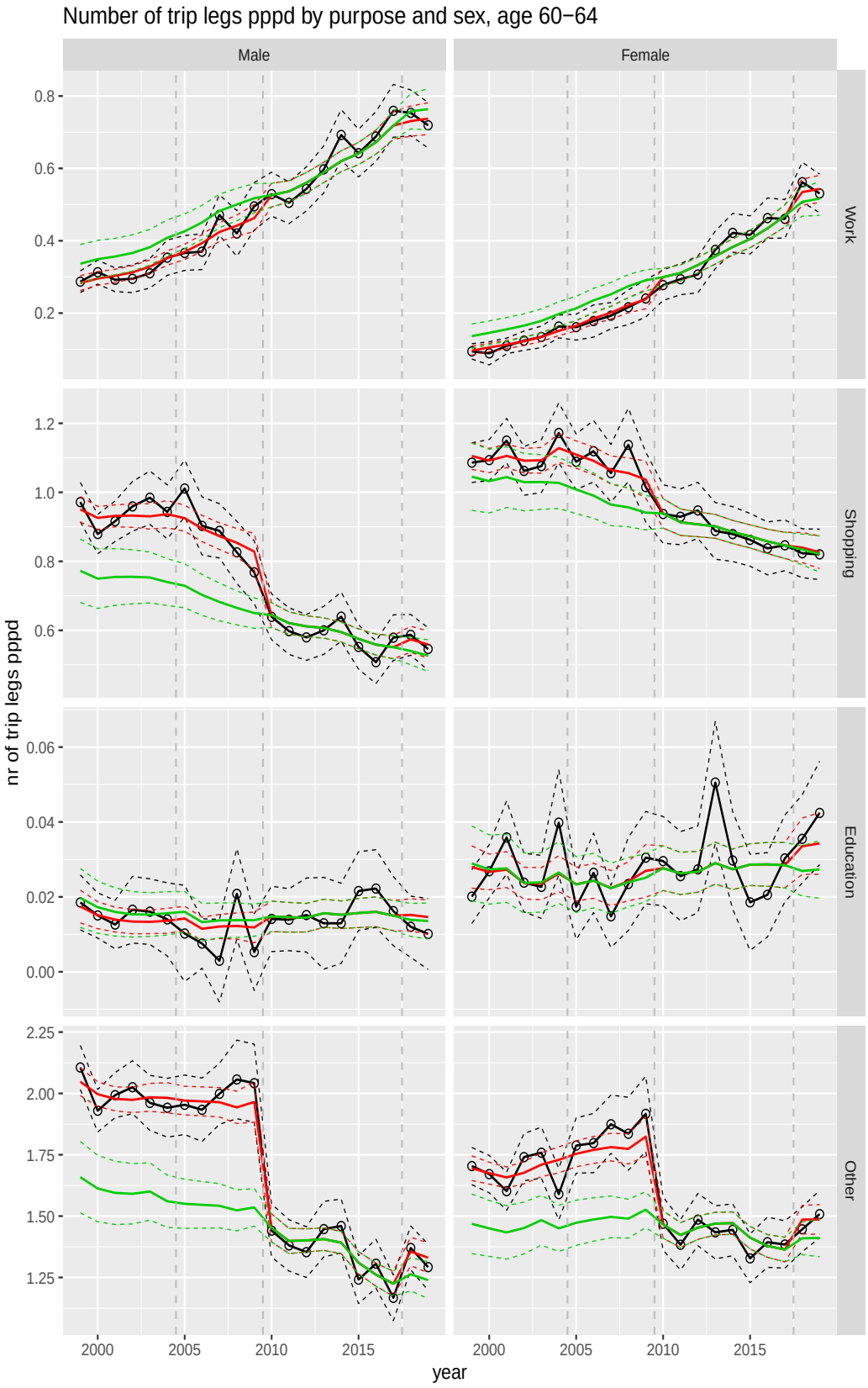

Figure A.14 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

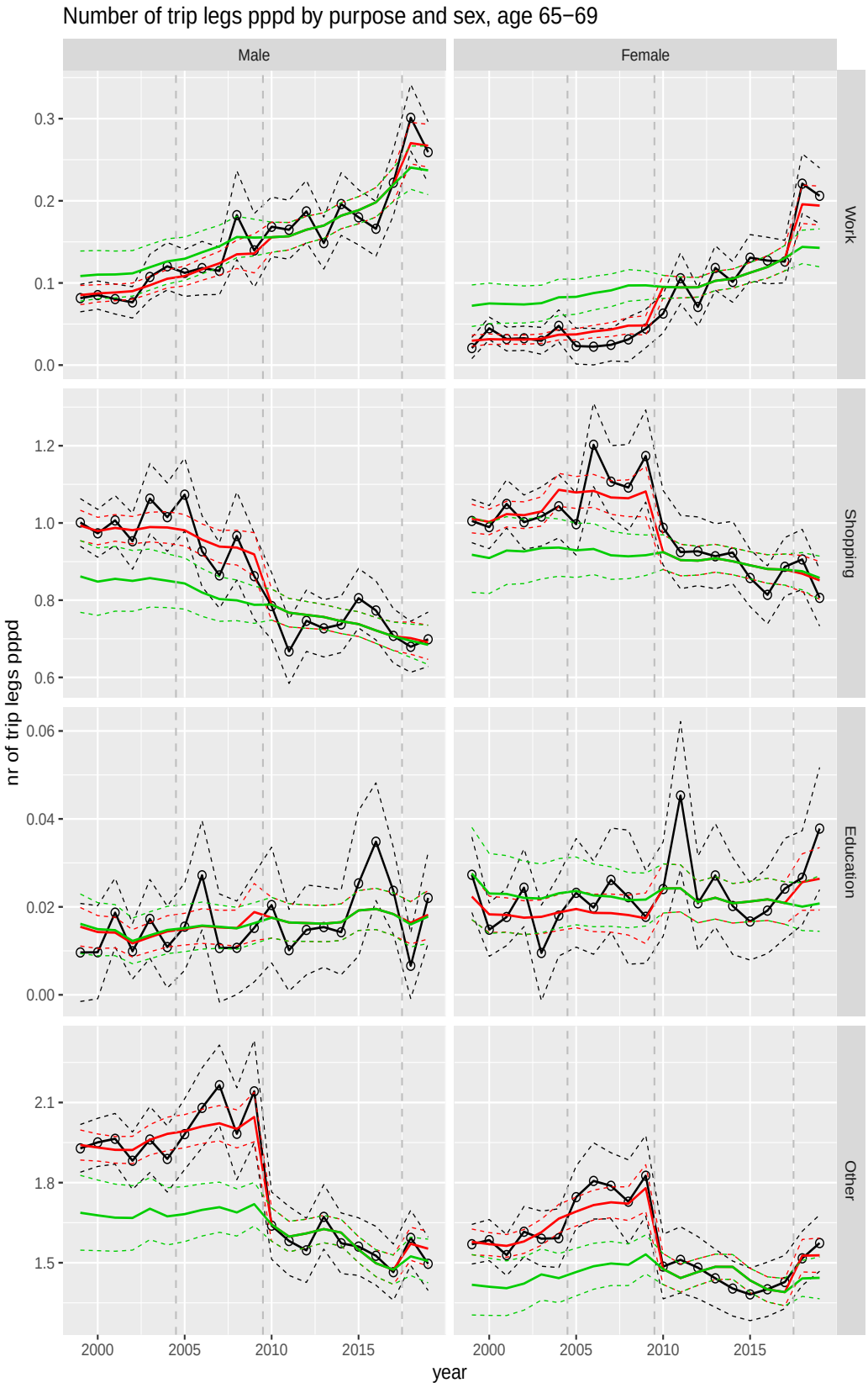

Figure A.15 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

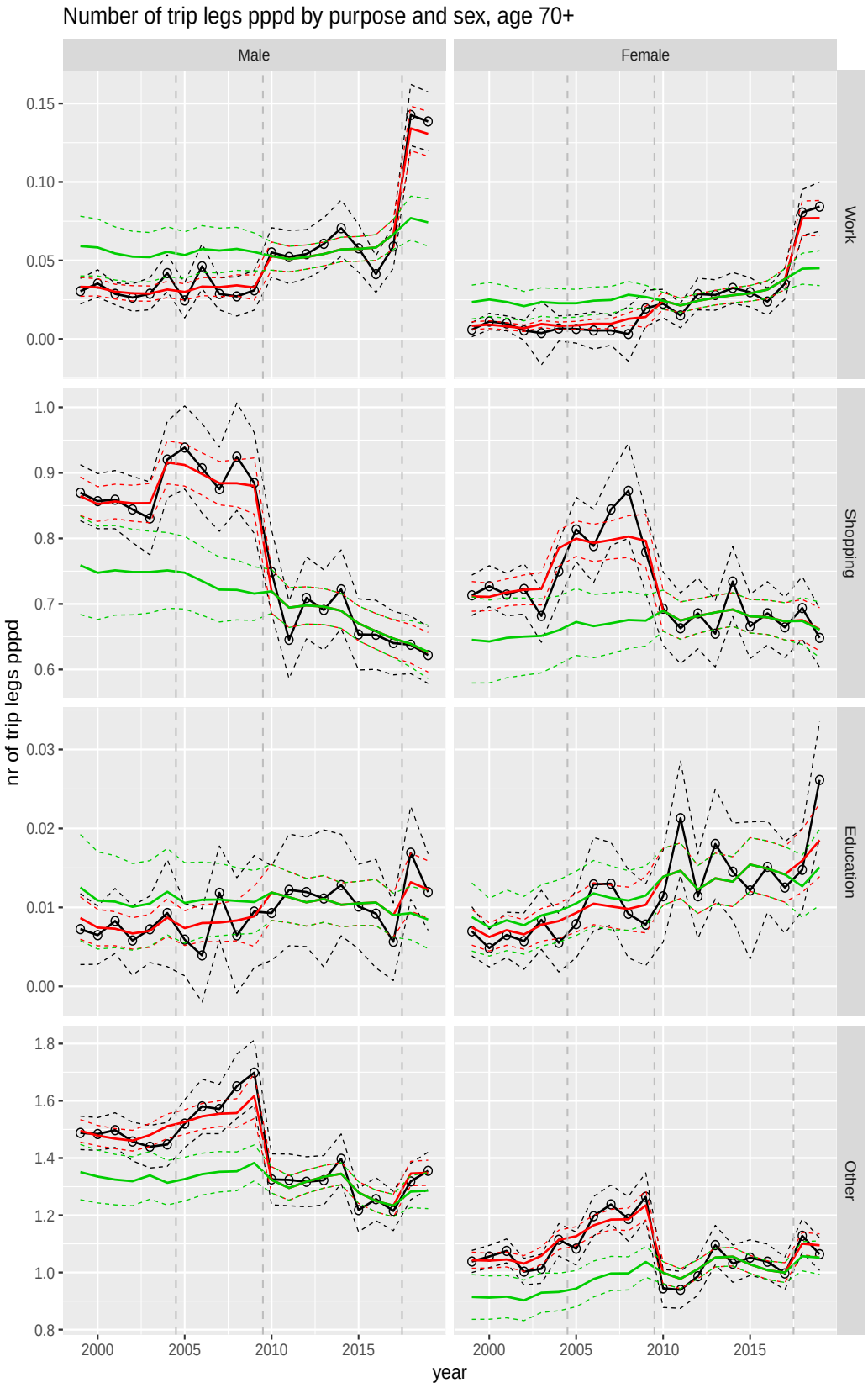

Figure A.16 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

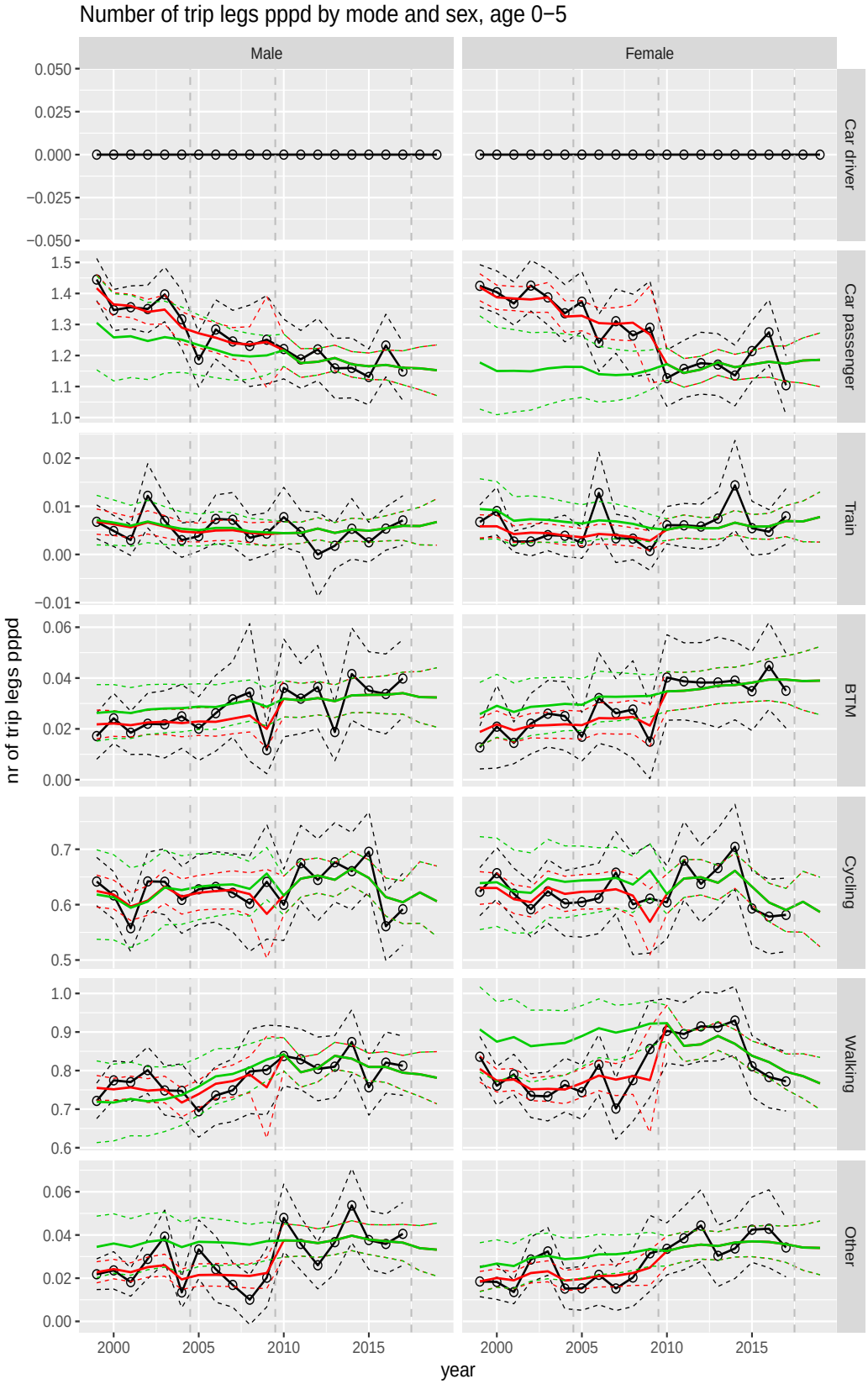

Figure A.17 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

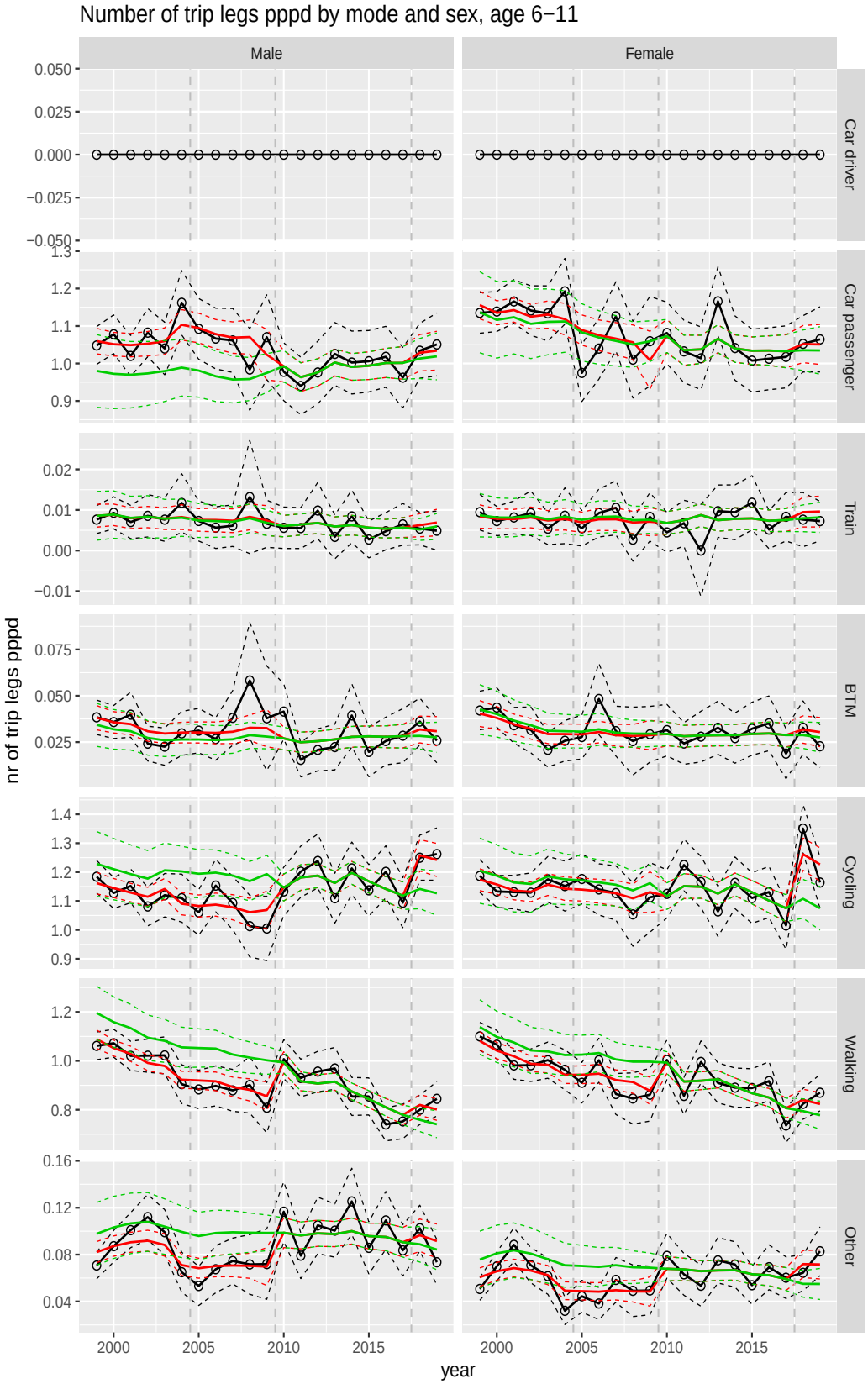

Figure A.18 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

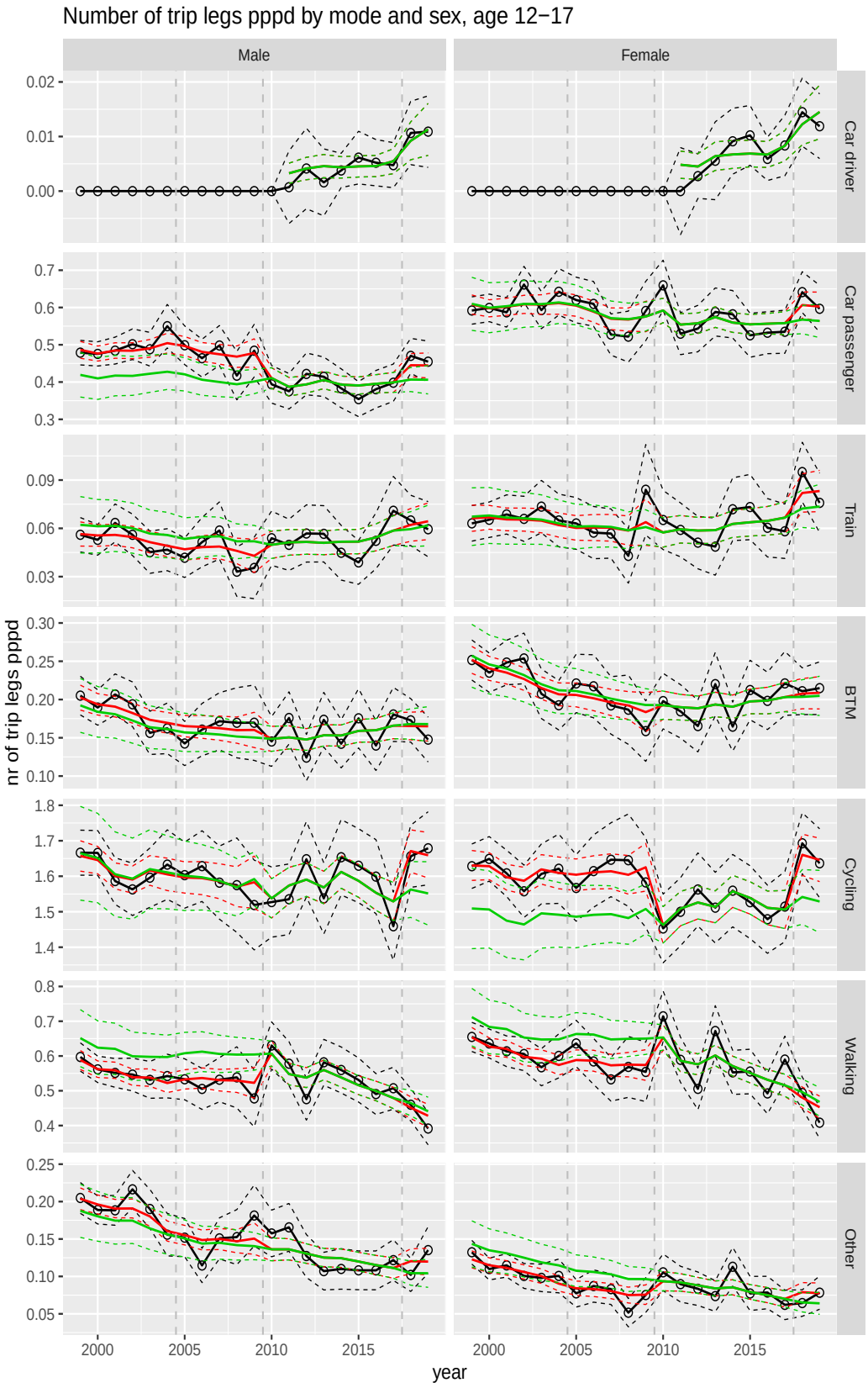

Figure A.19 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

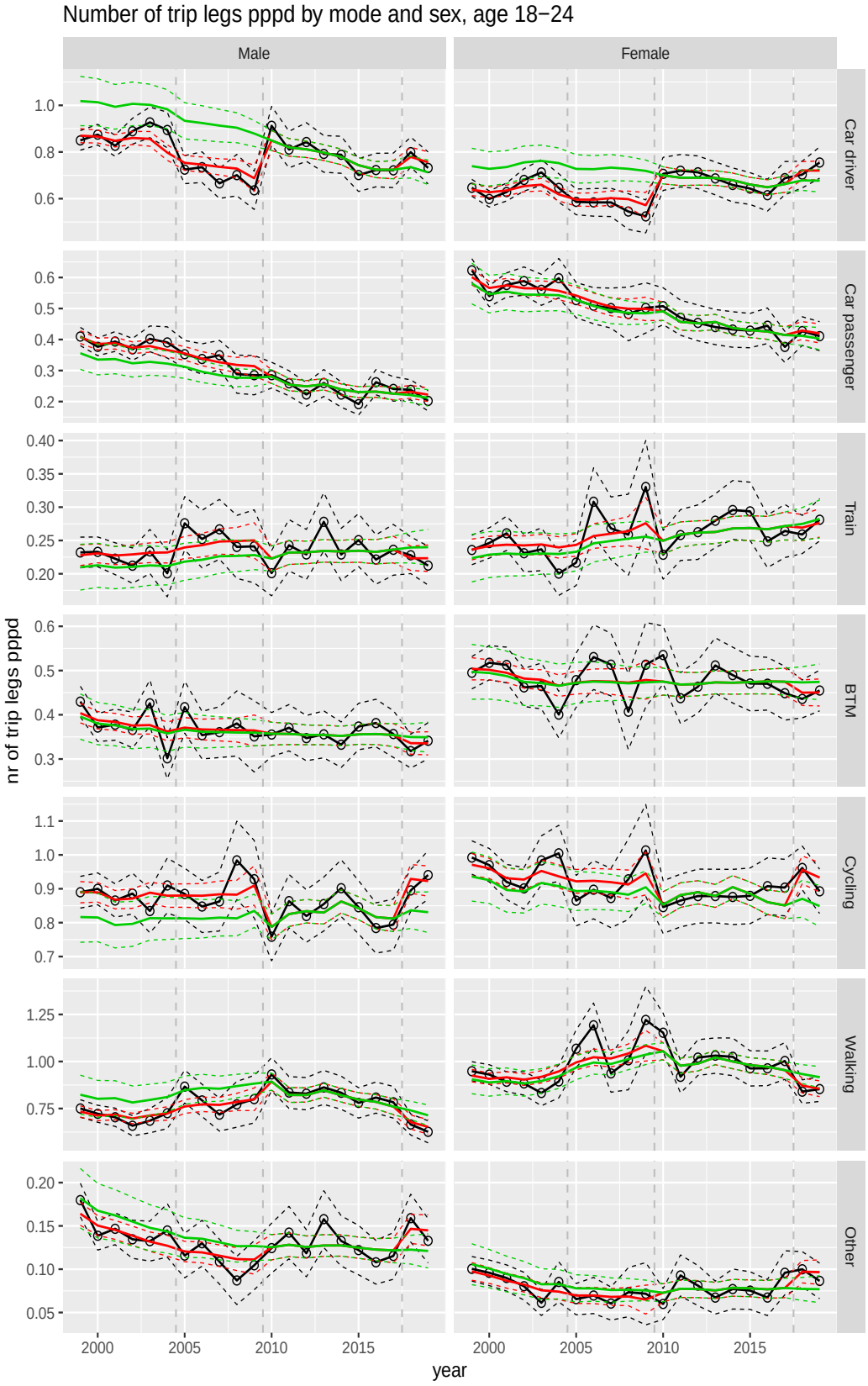

Figure A.20 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

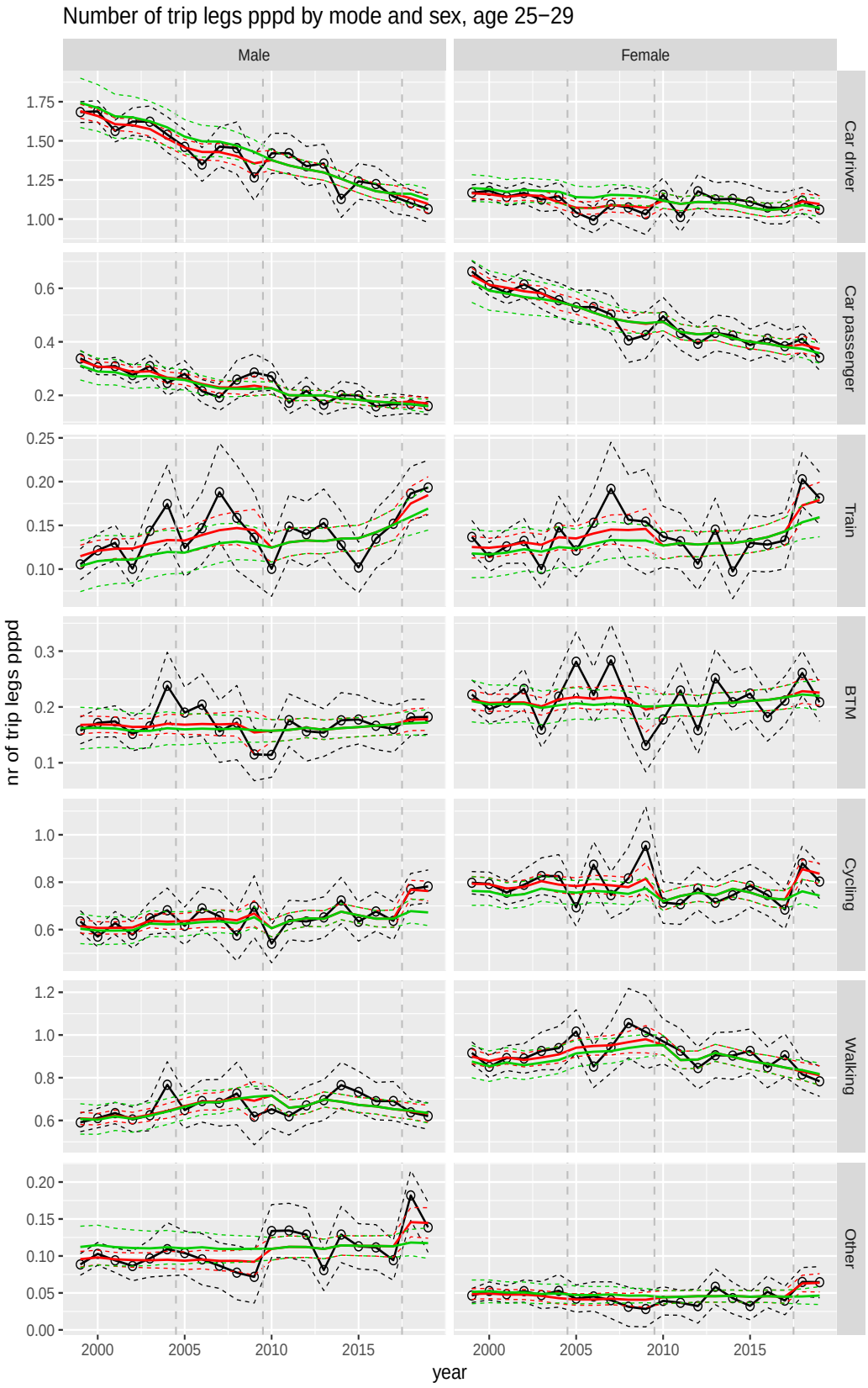

Figure A.21 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

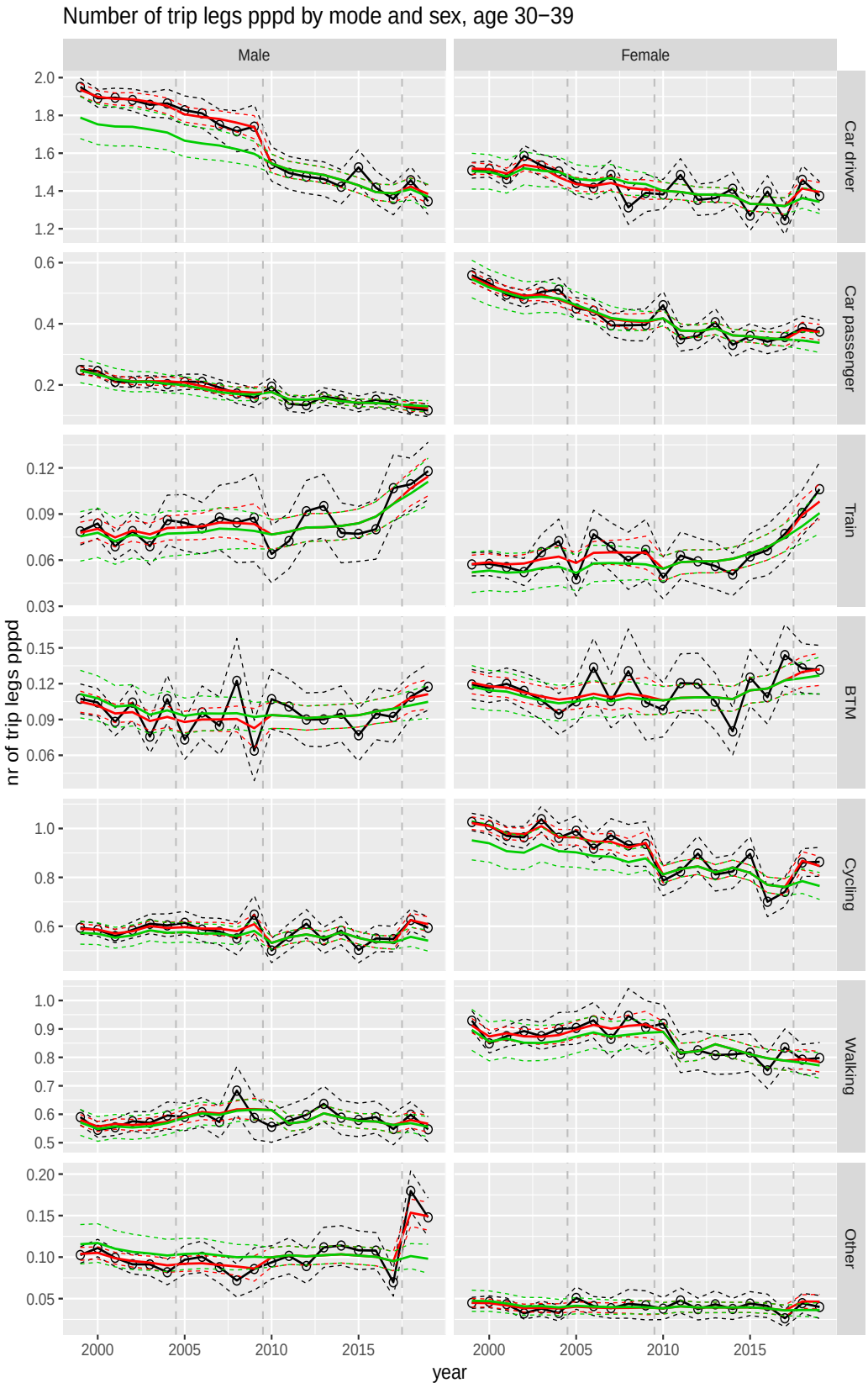

Figure A.22 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

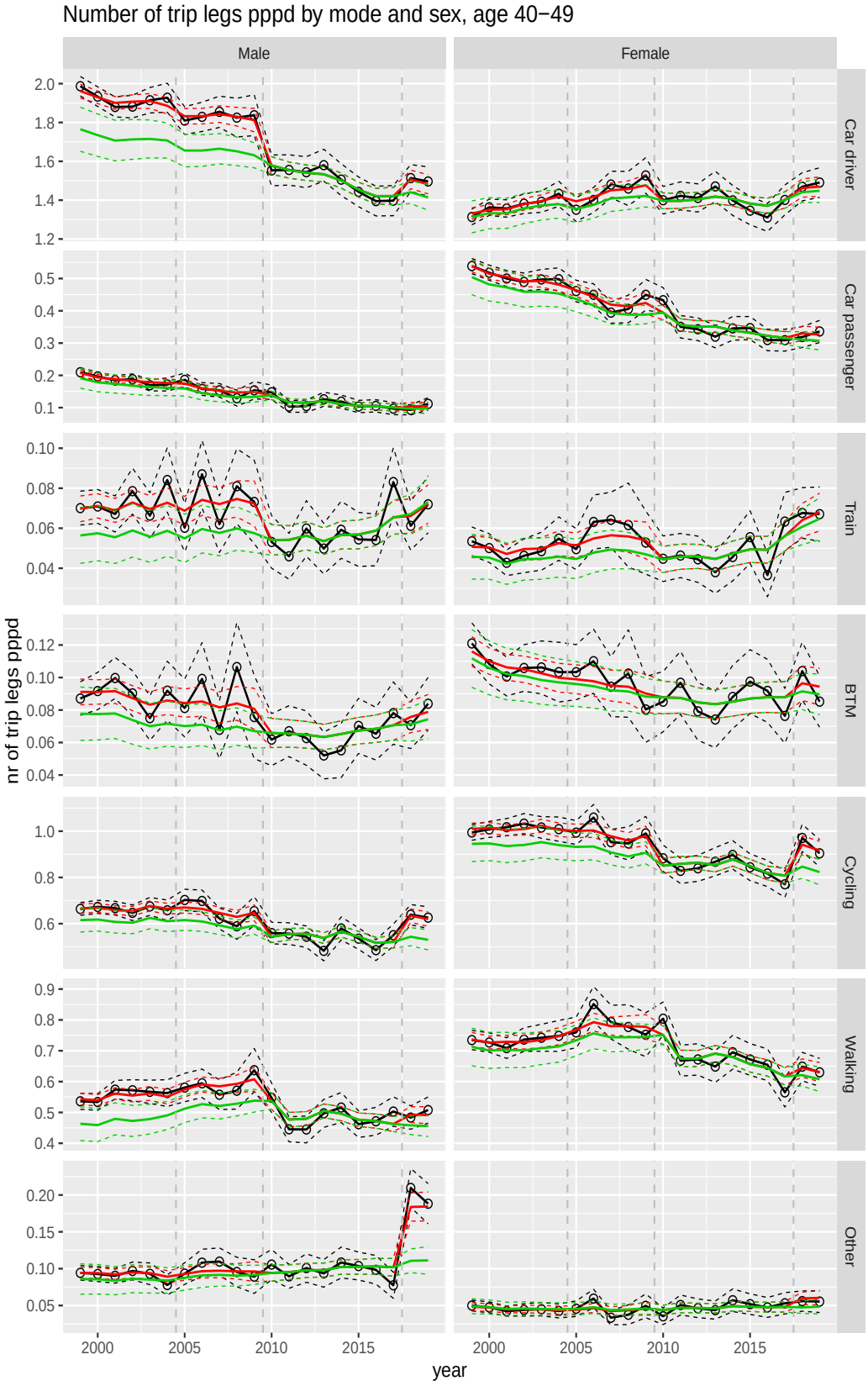

Figure A.23 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

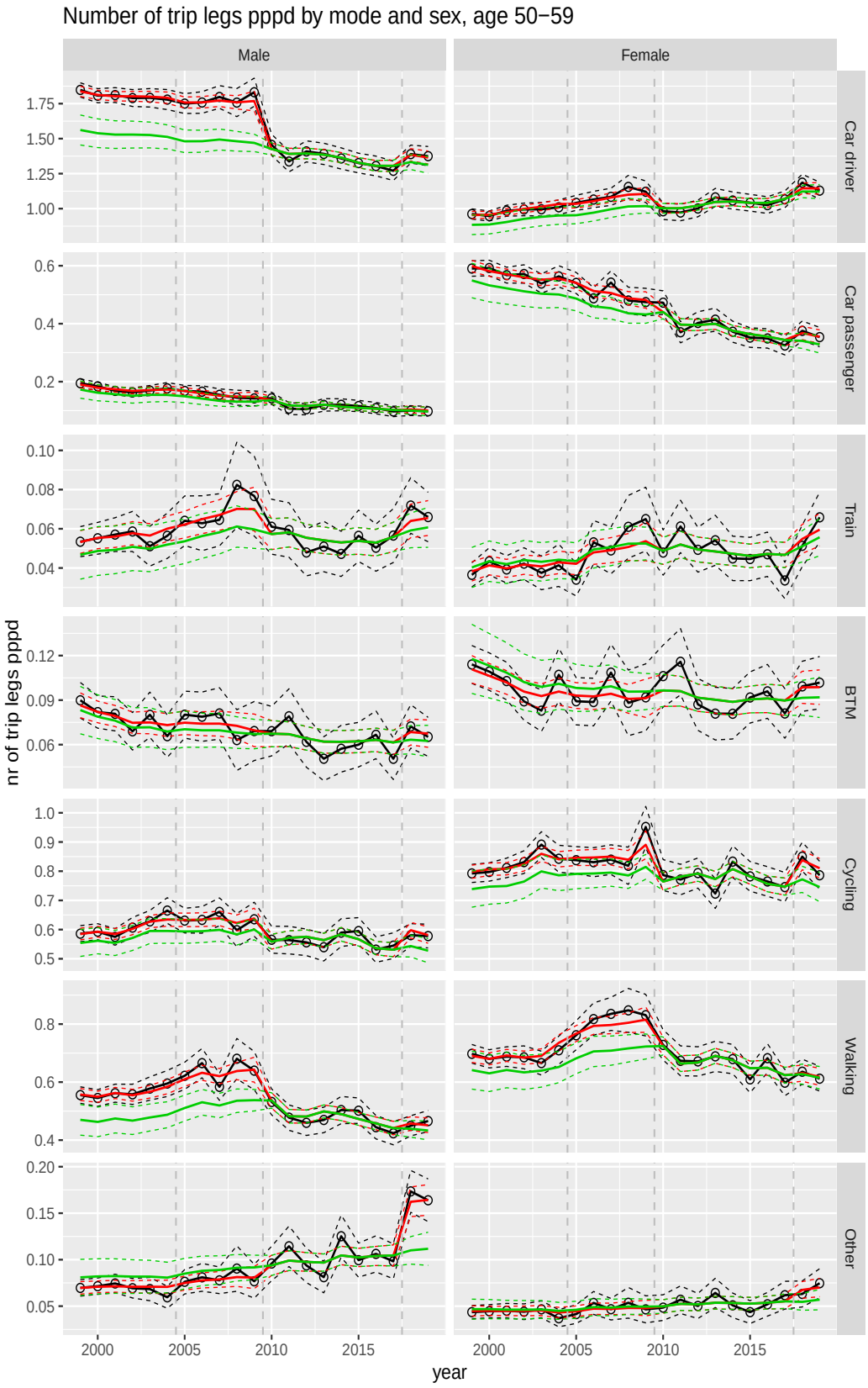

Figure A.24 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

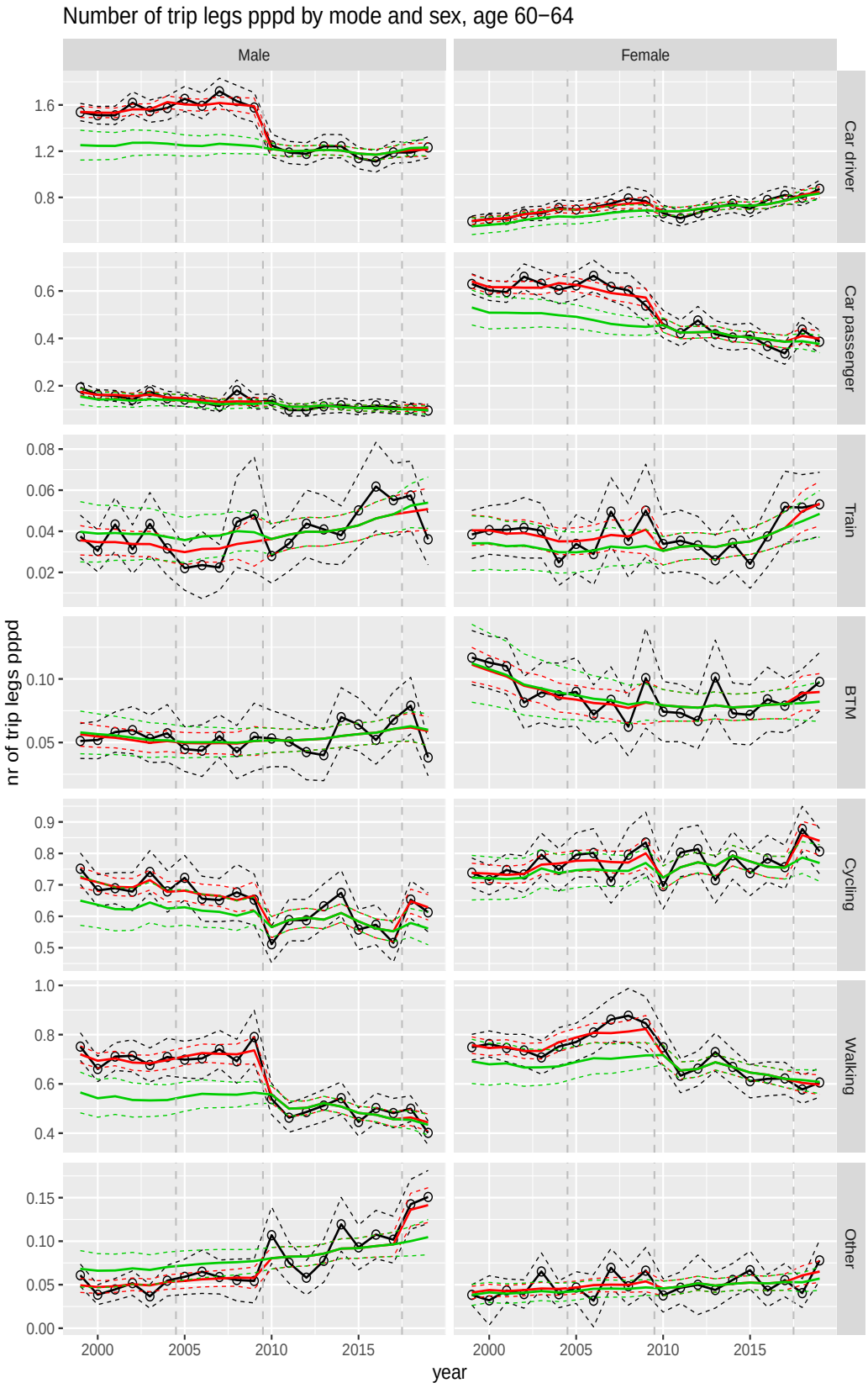

Figure A.25 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

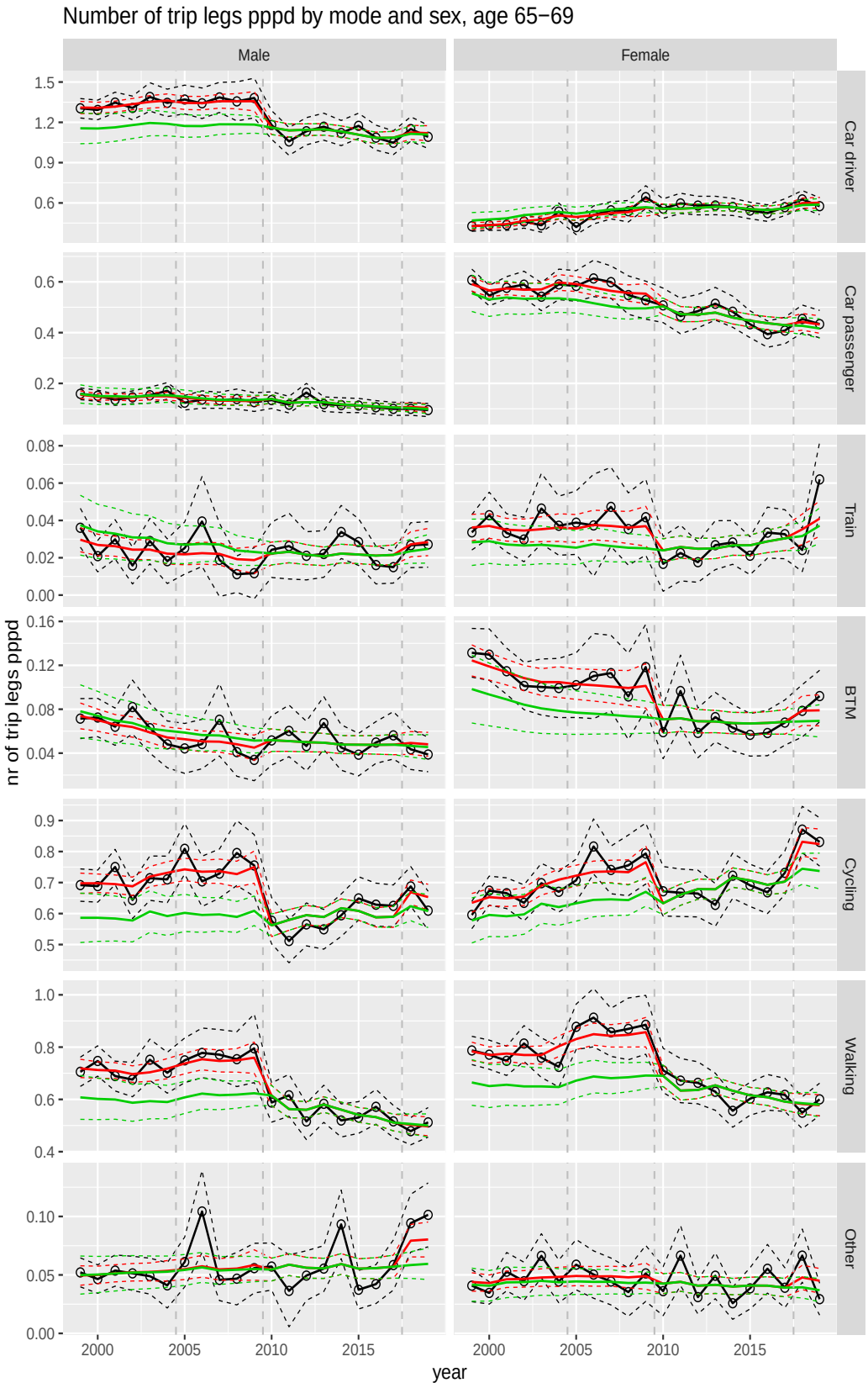

Figure A.26 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

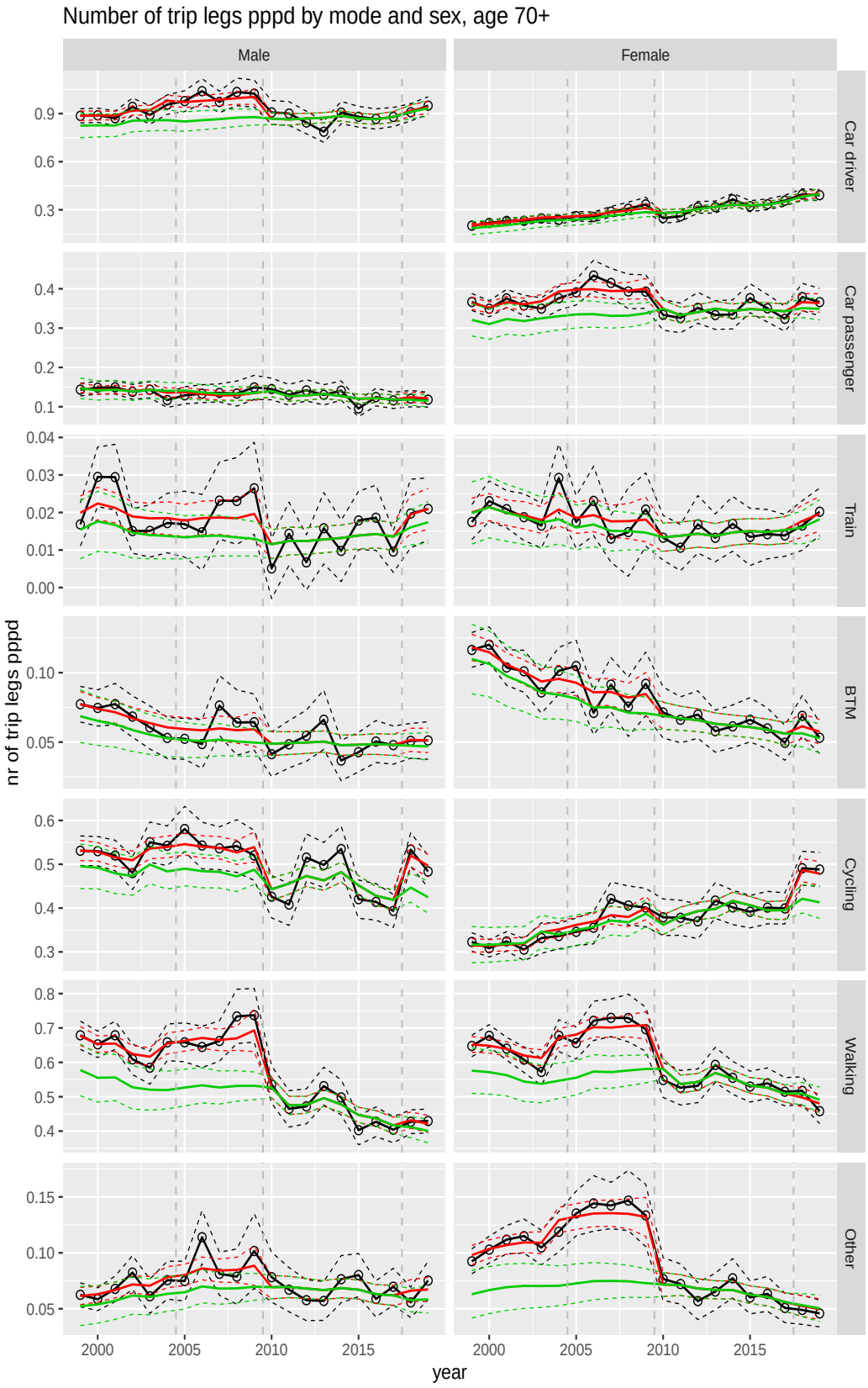

Figure A.27 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

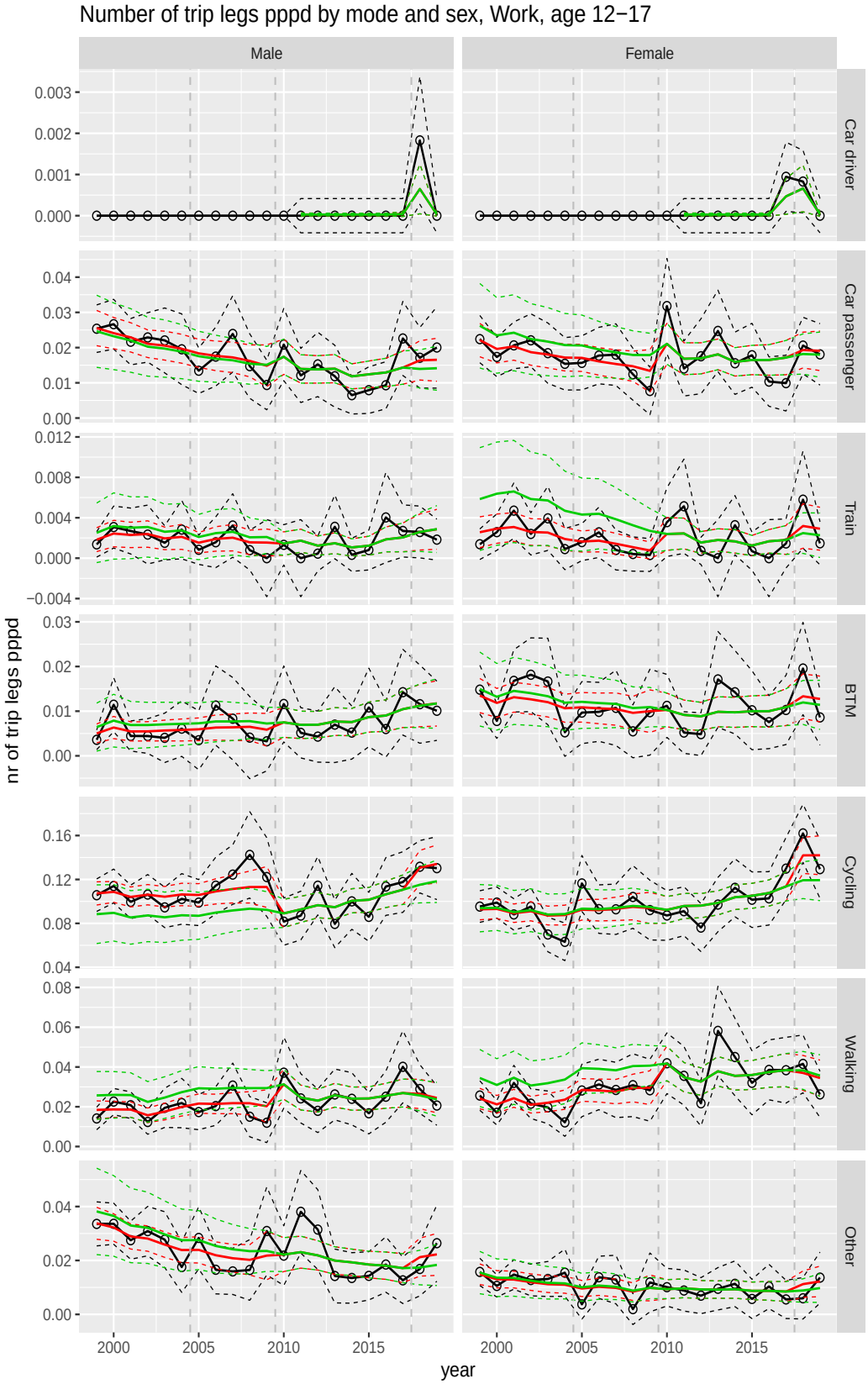

Figure A.28 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

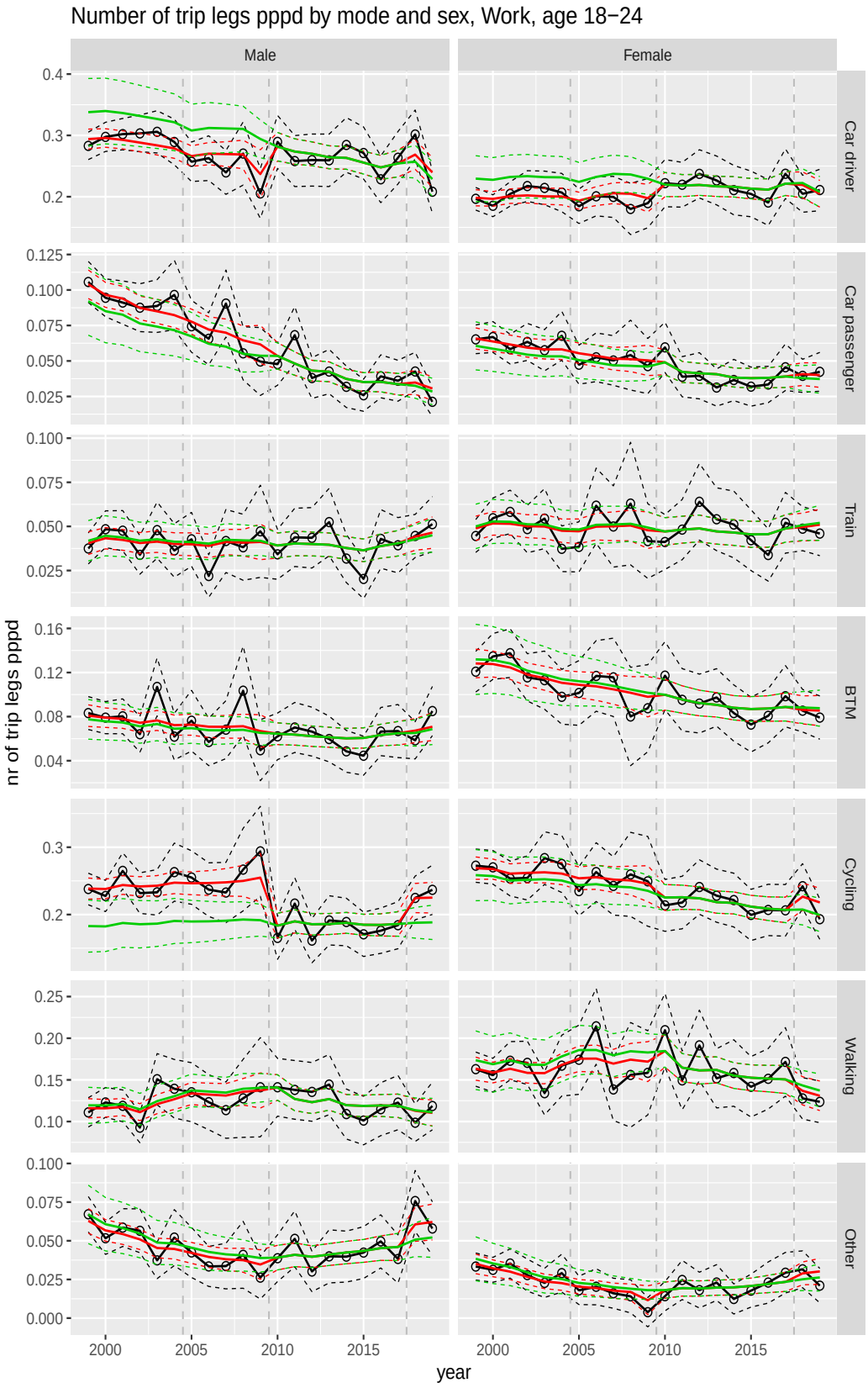

Figure A.29 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

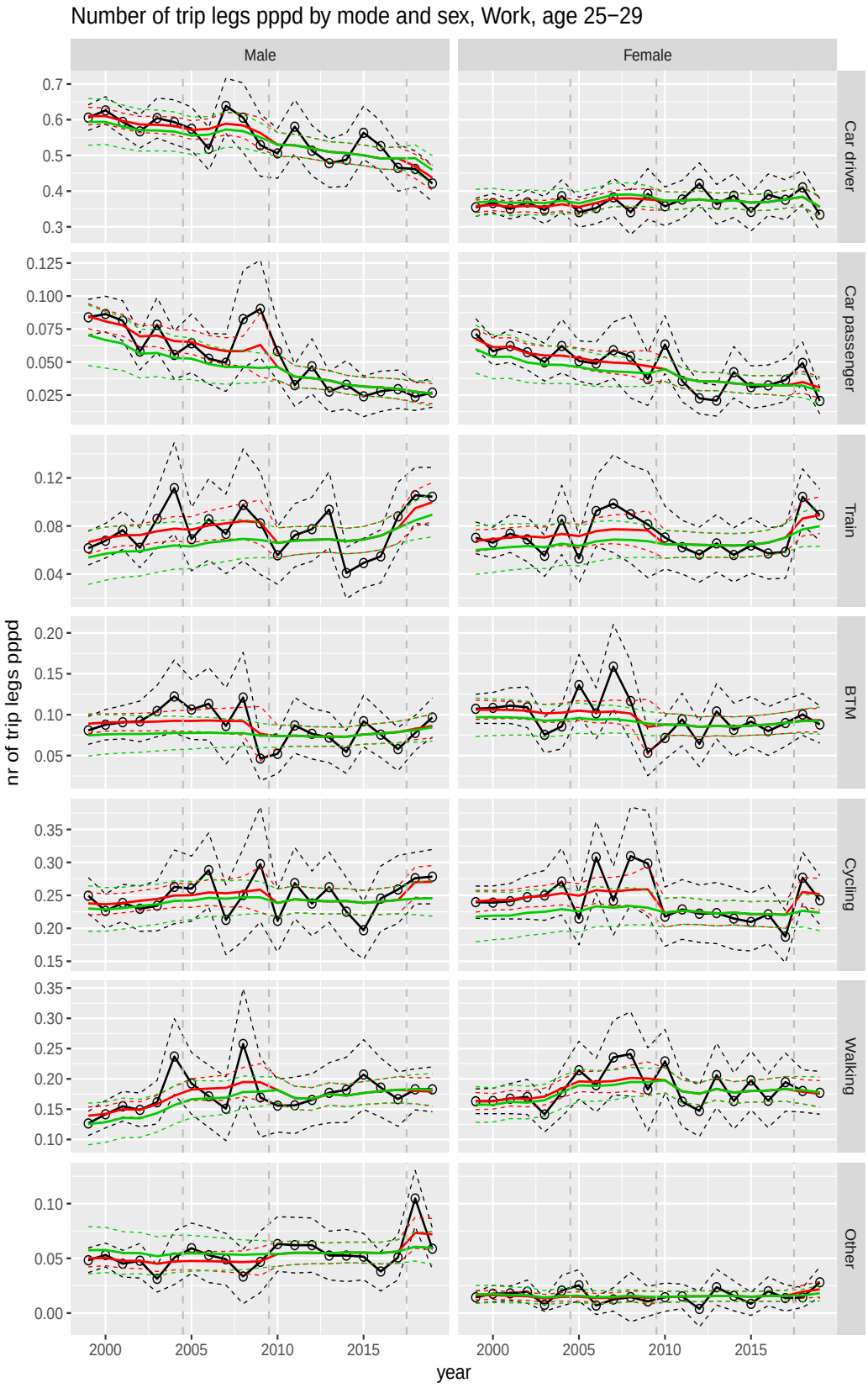

Figure A.30 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure A.31 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

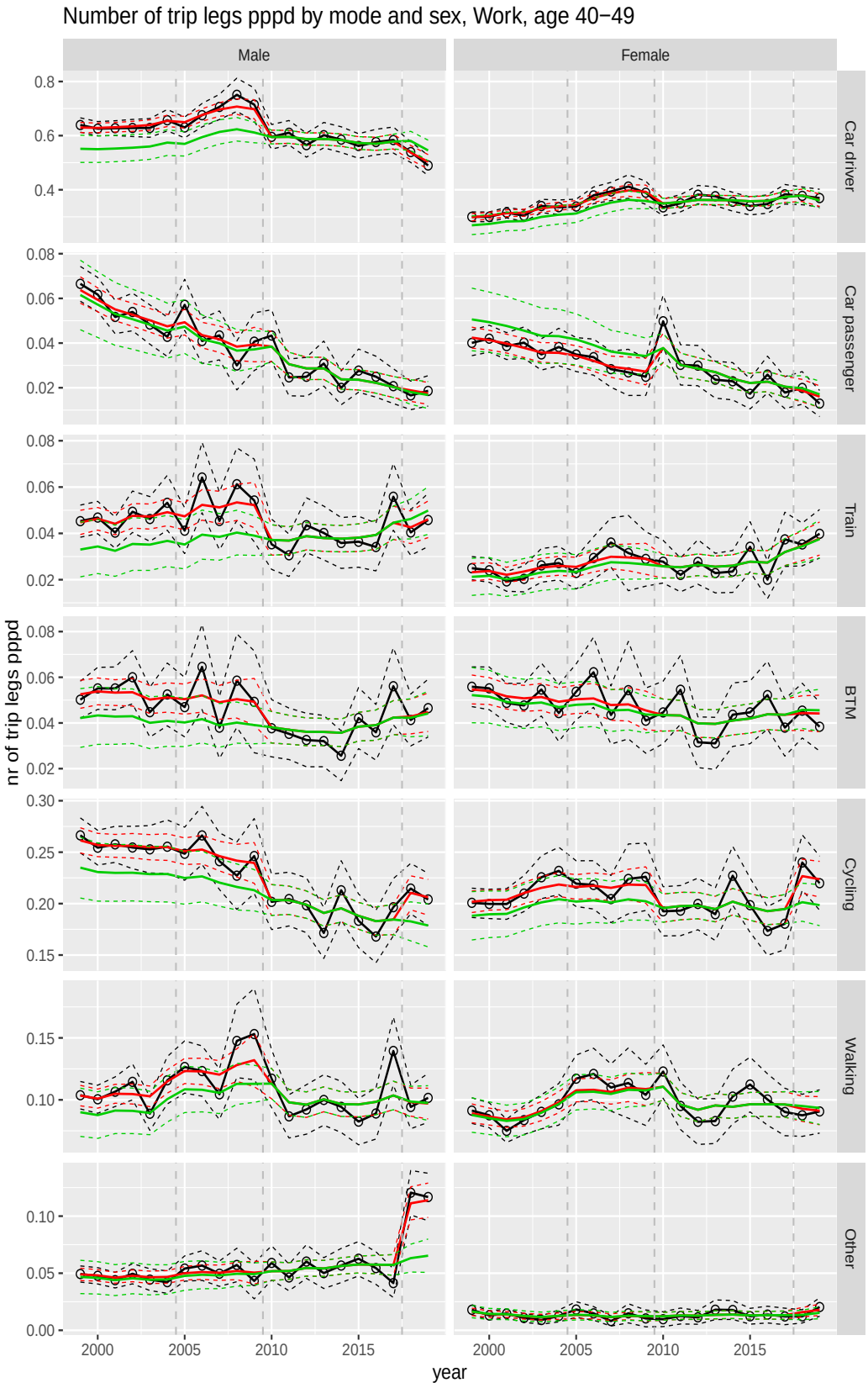

Figure A.32 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

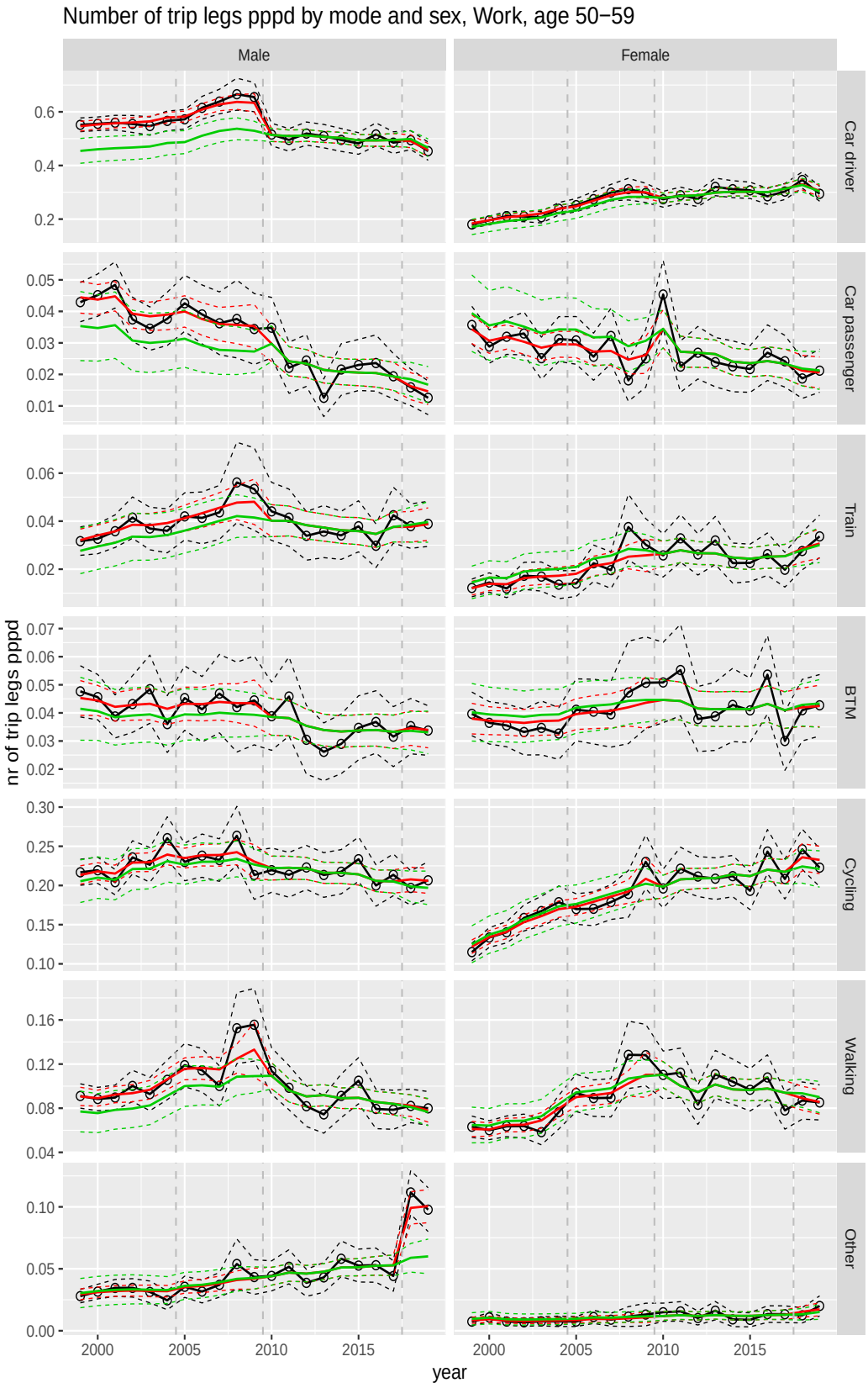

Figure A.33 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

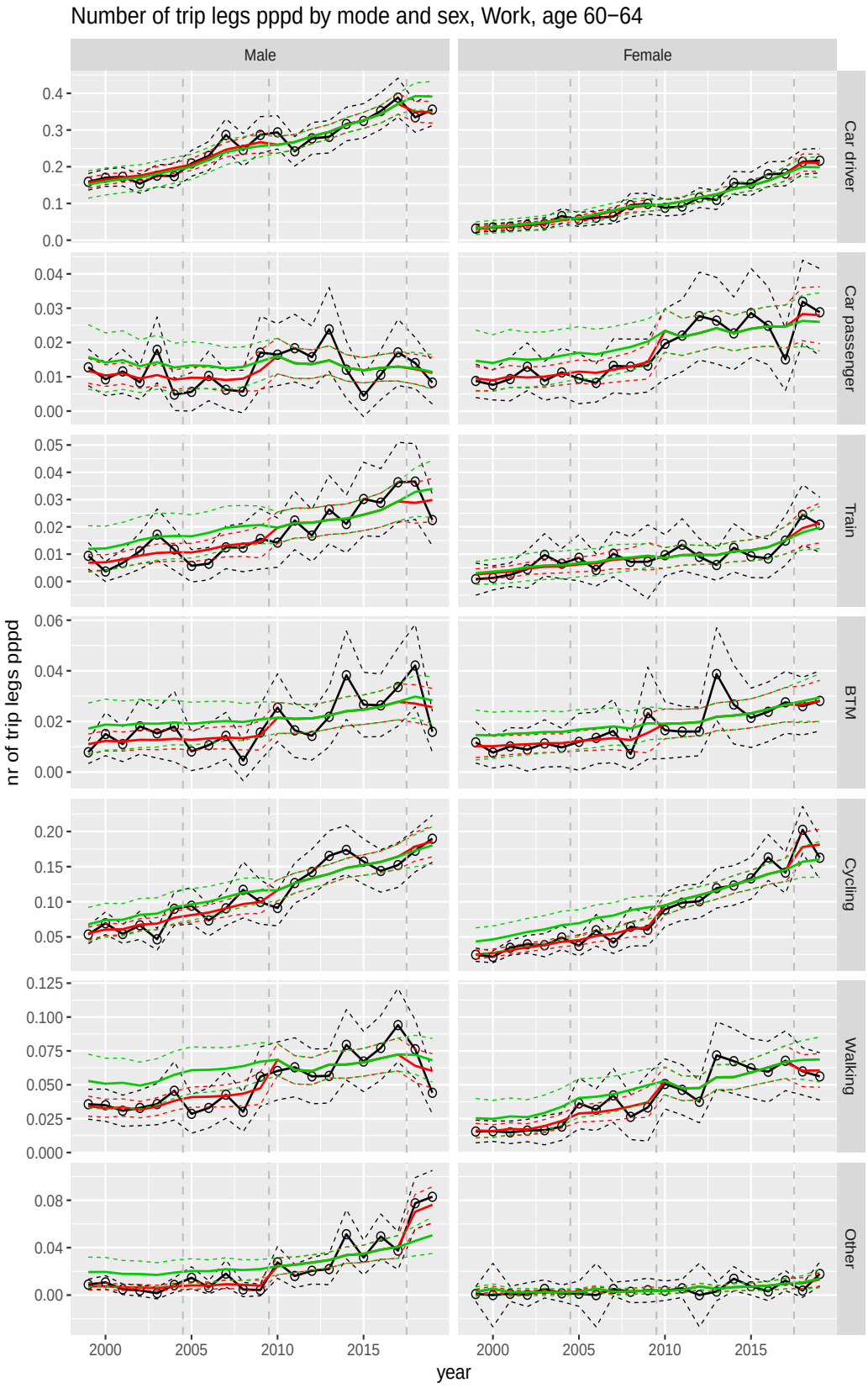

Figure A.34 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

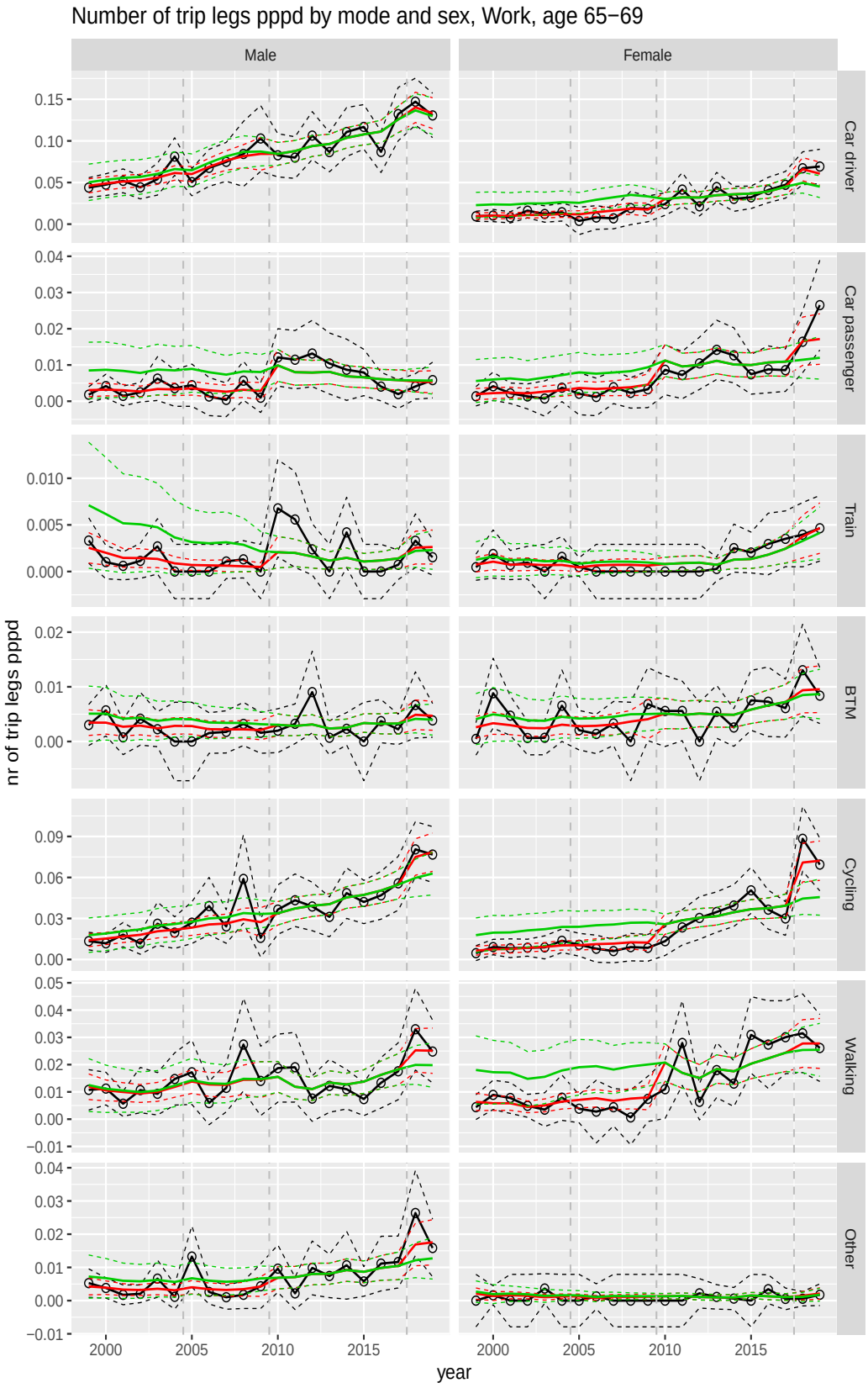

Figure A.35 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

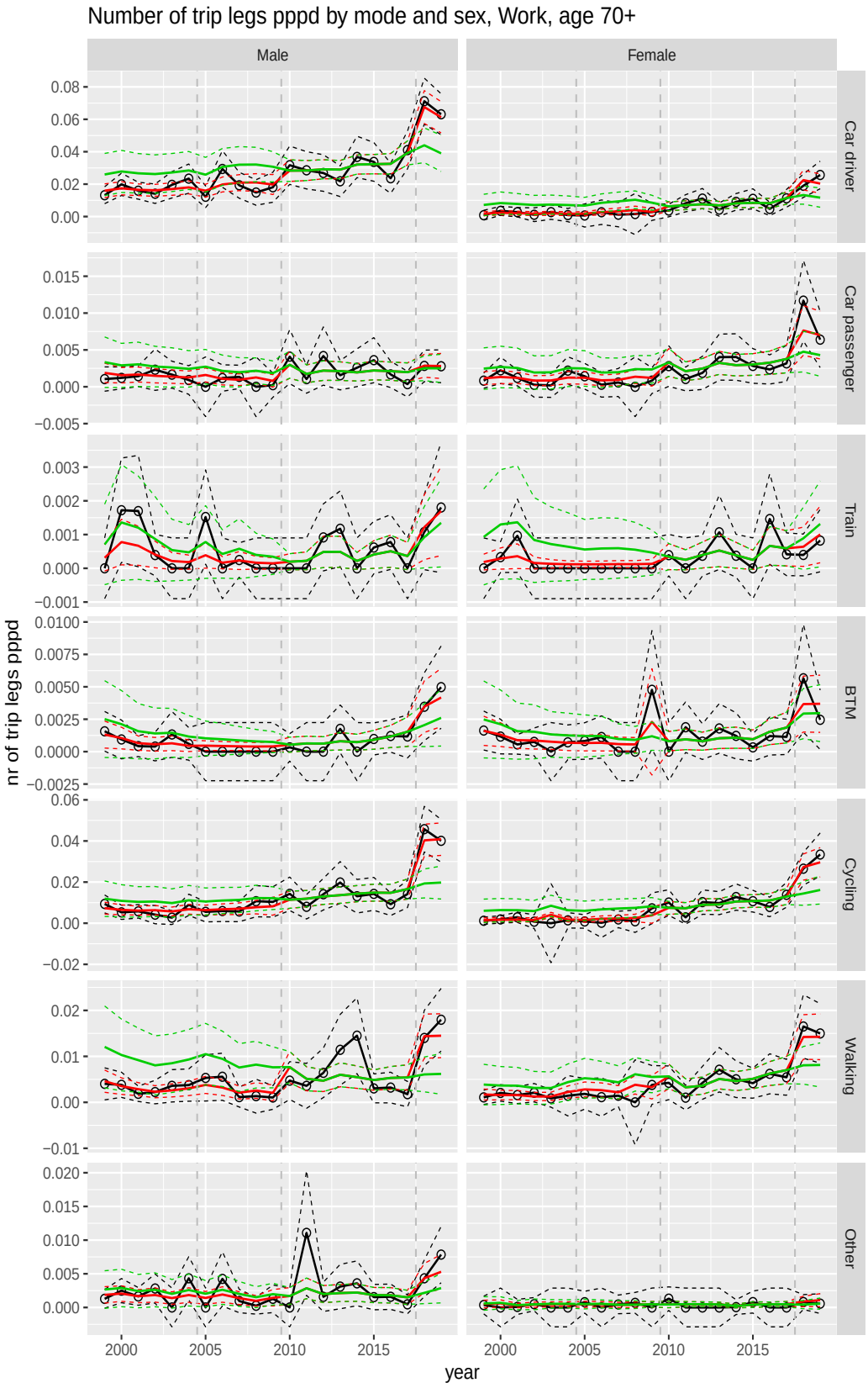

Figure A.36 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

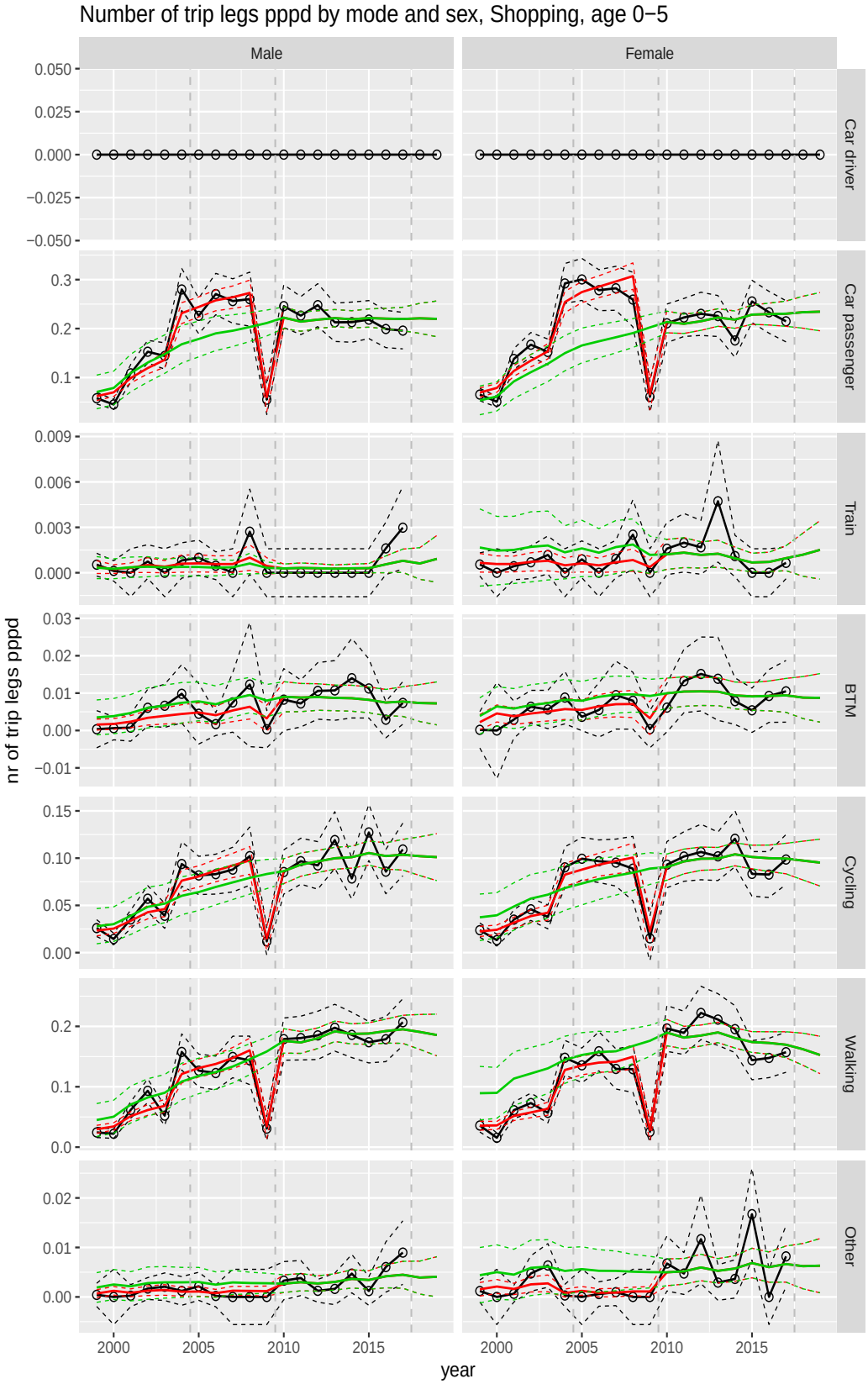

Figure A.37 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

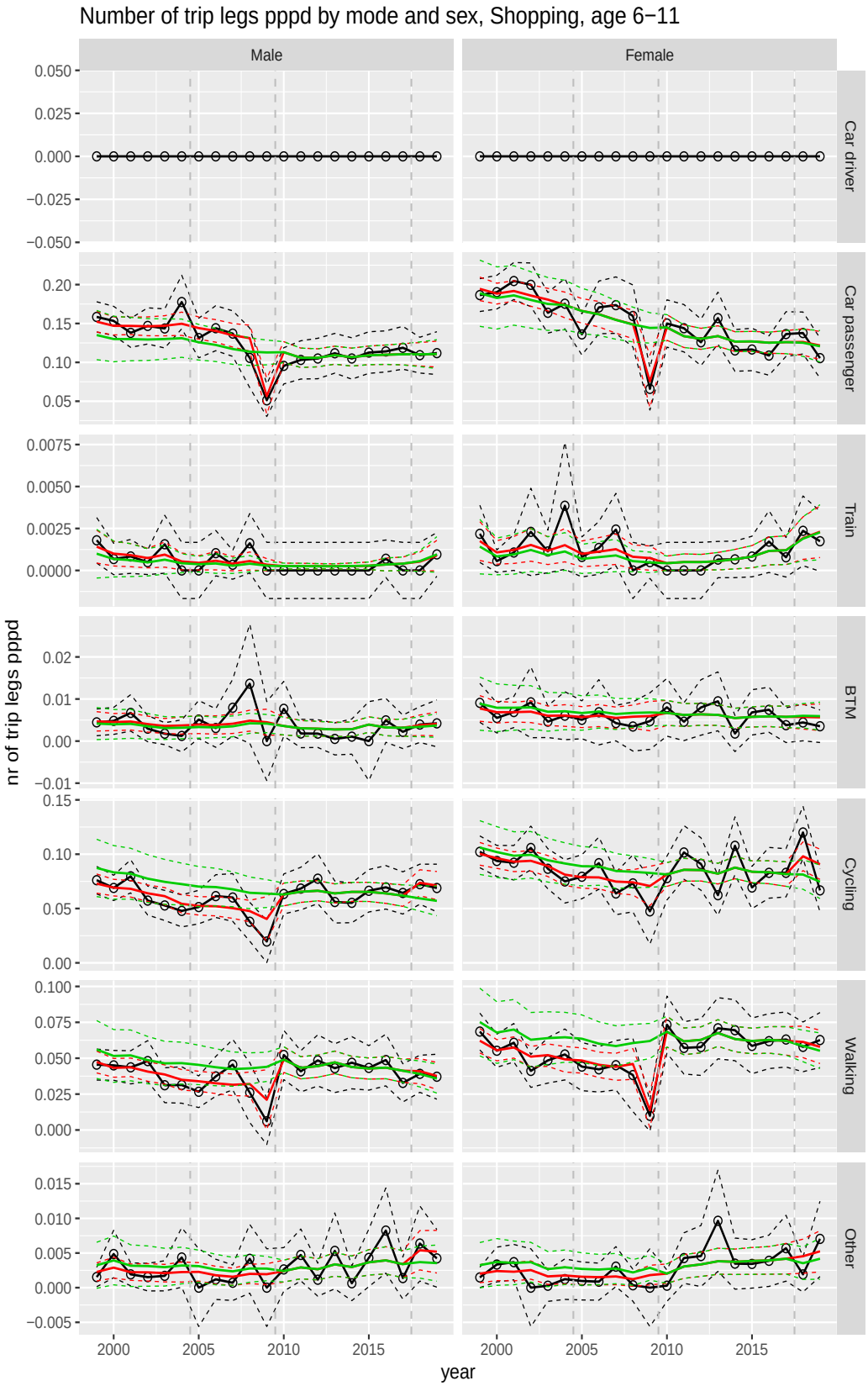

Figure A.38 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

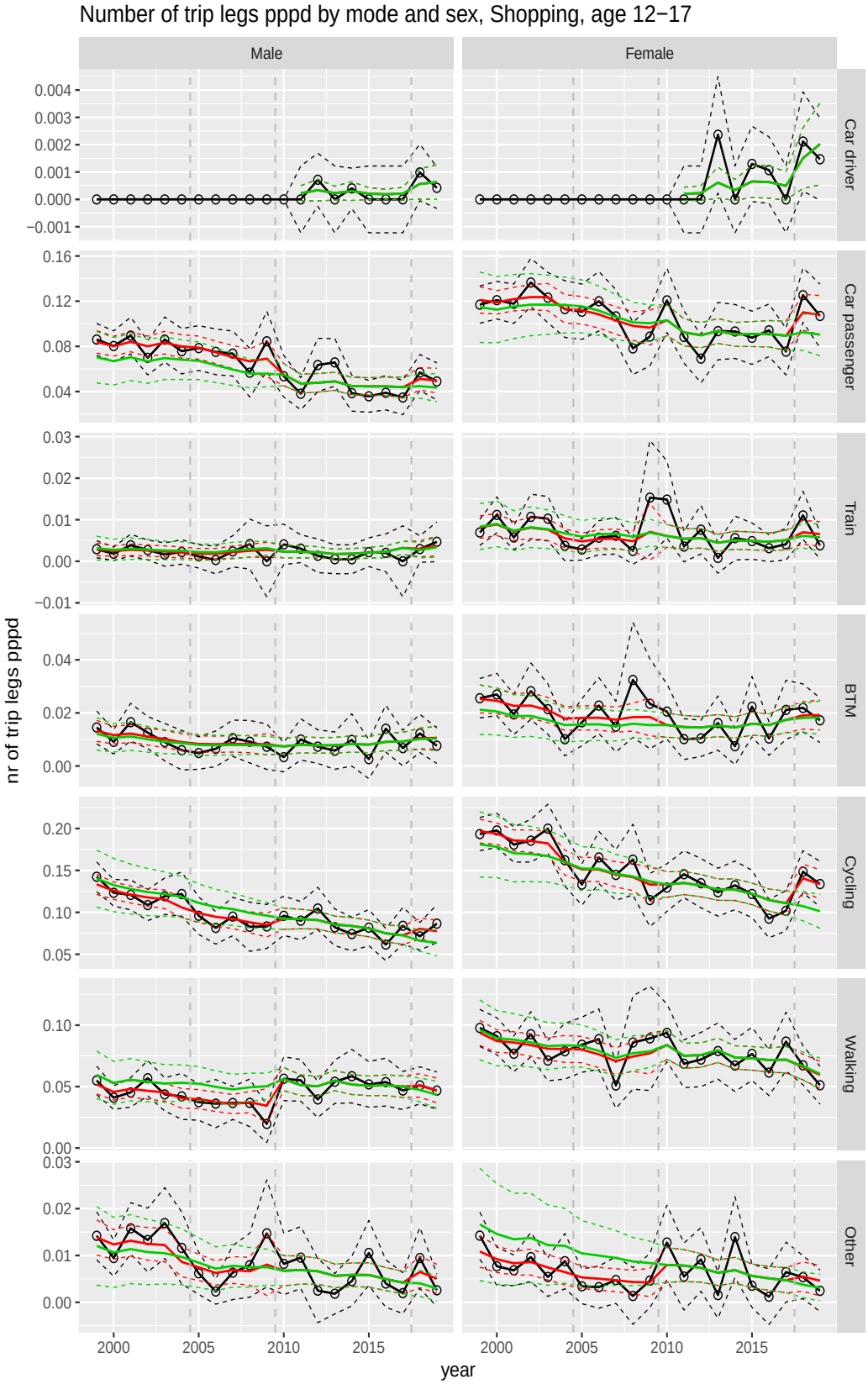

Figure A.39 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

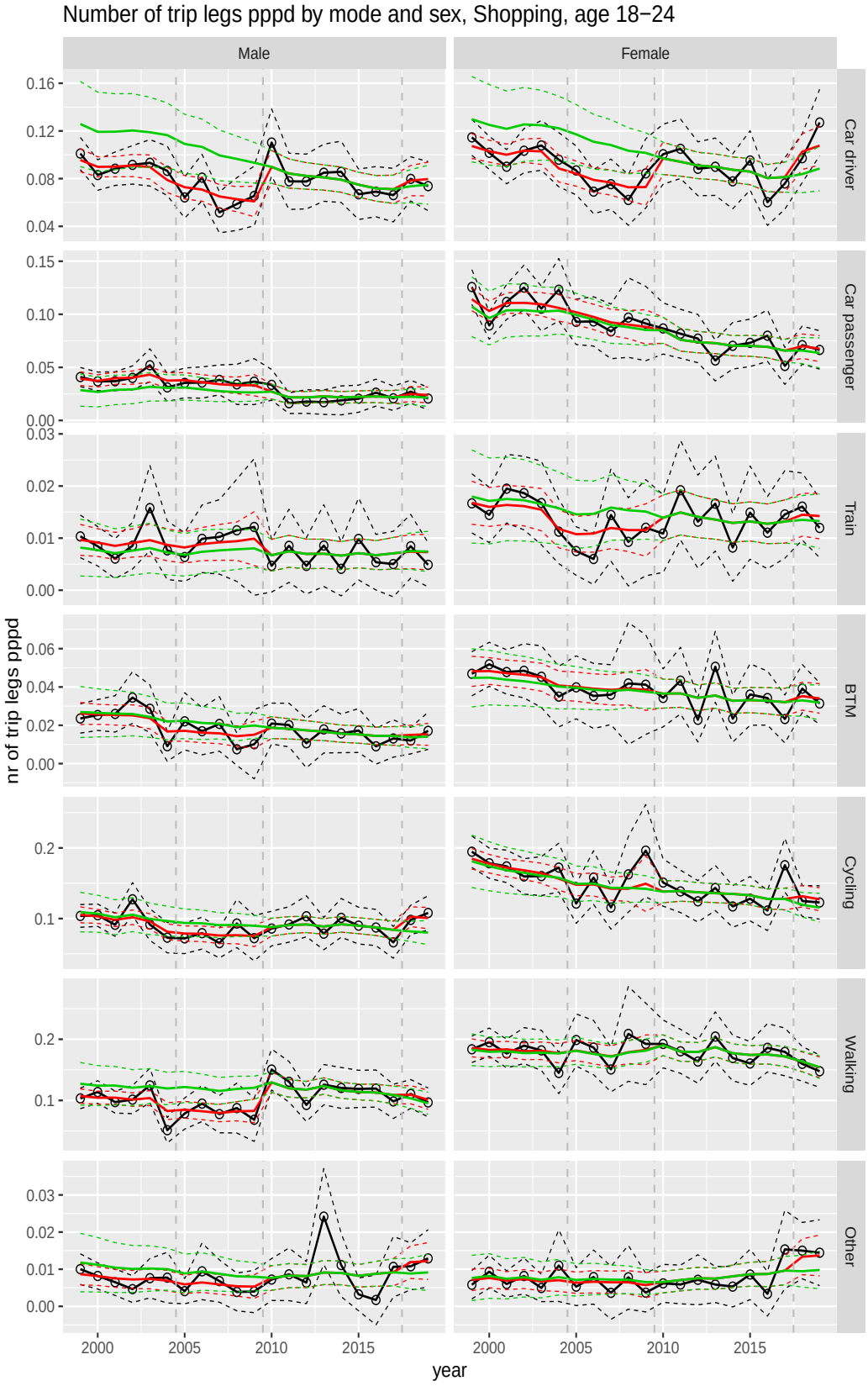

Figure A.40 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

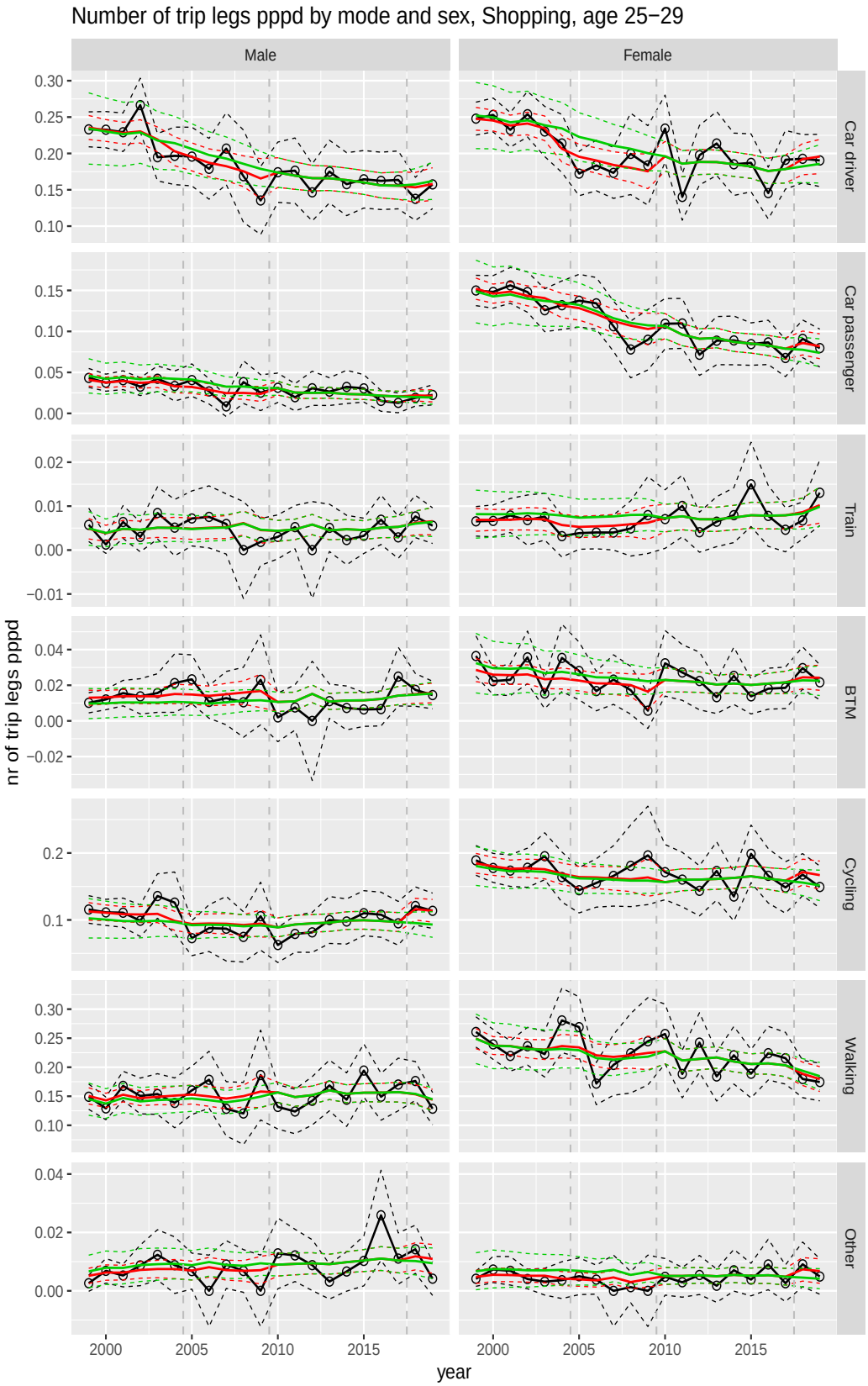

Figure A.41 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

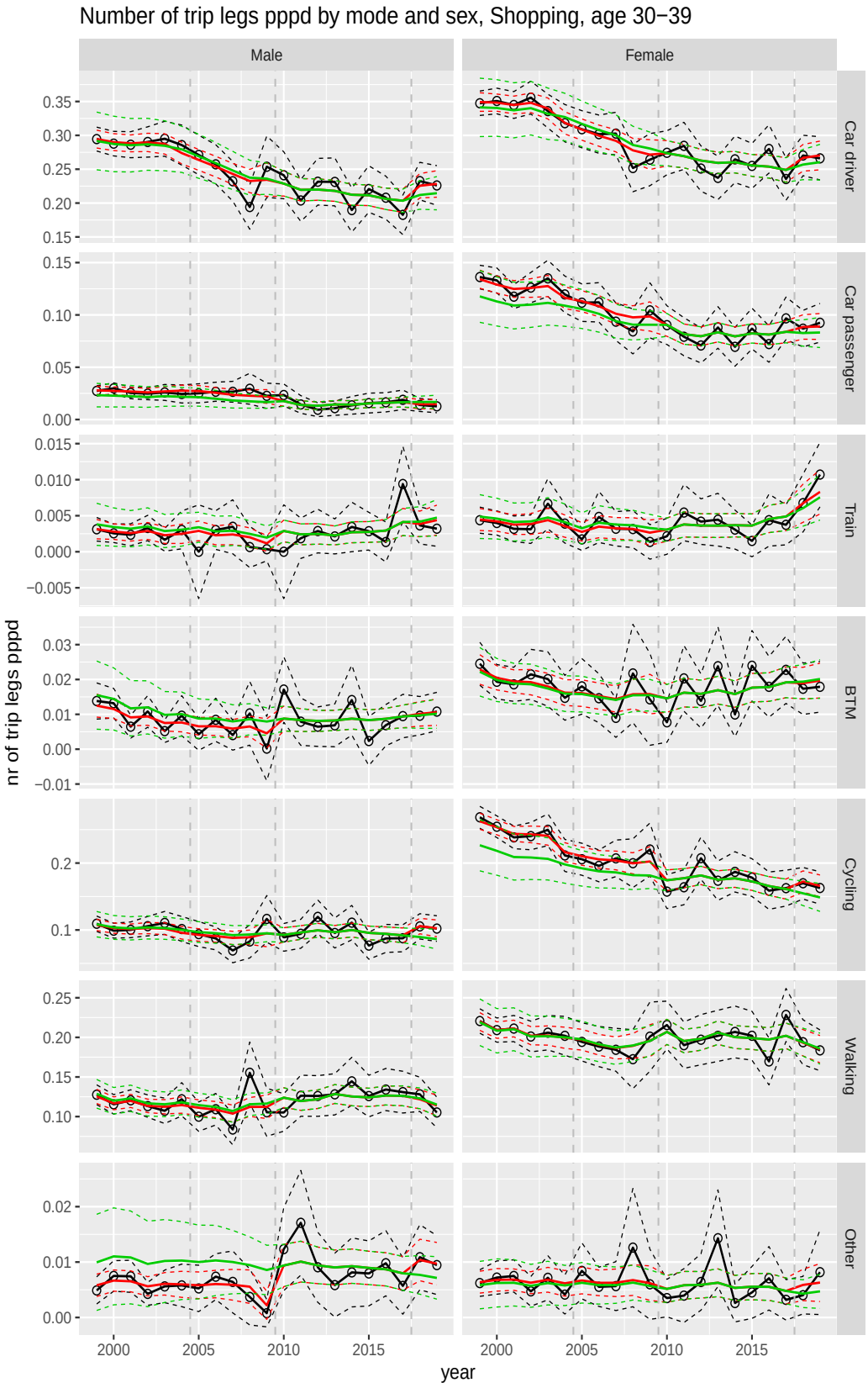

Figure A.42 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

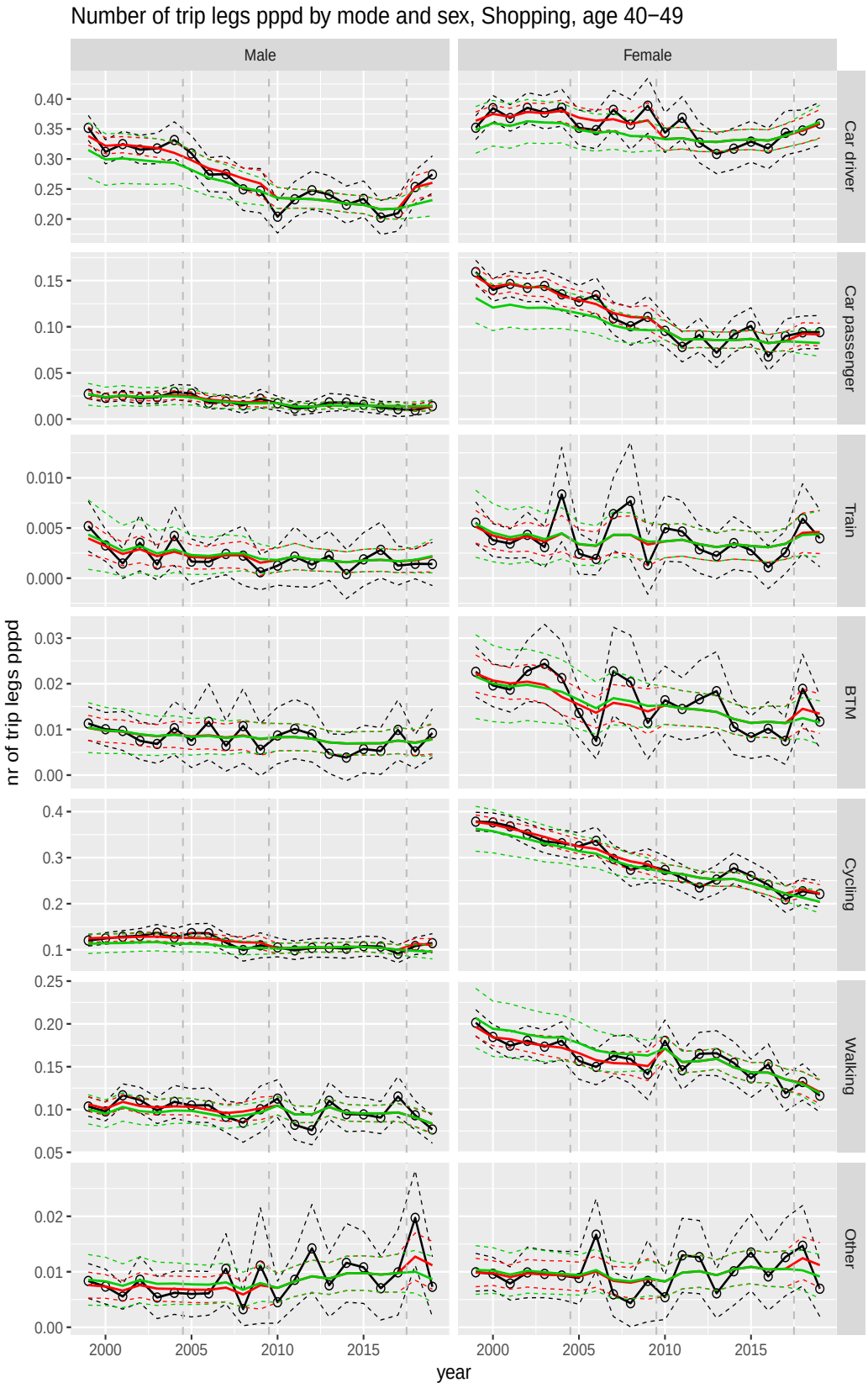

Figure A.43 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

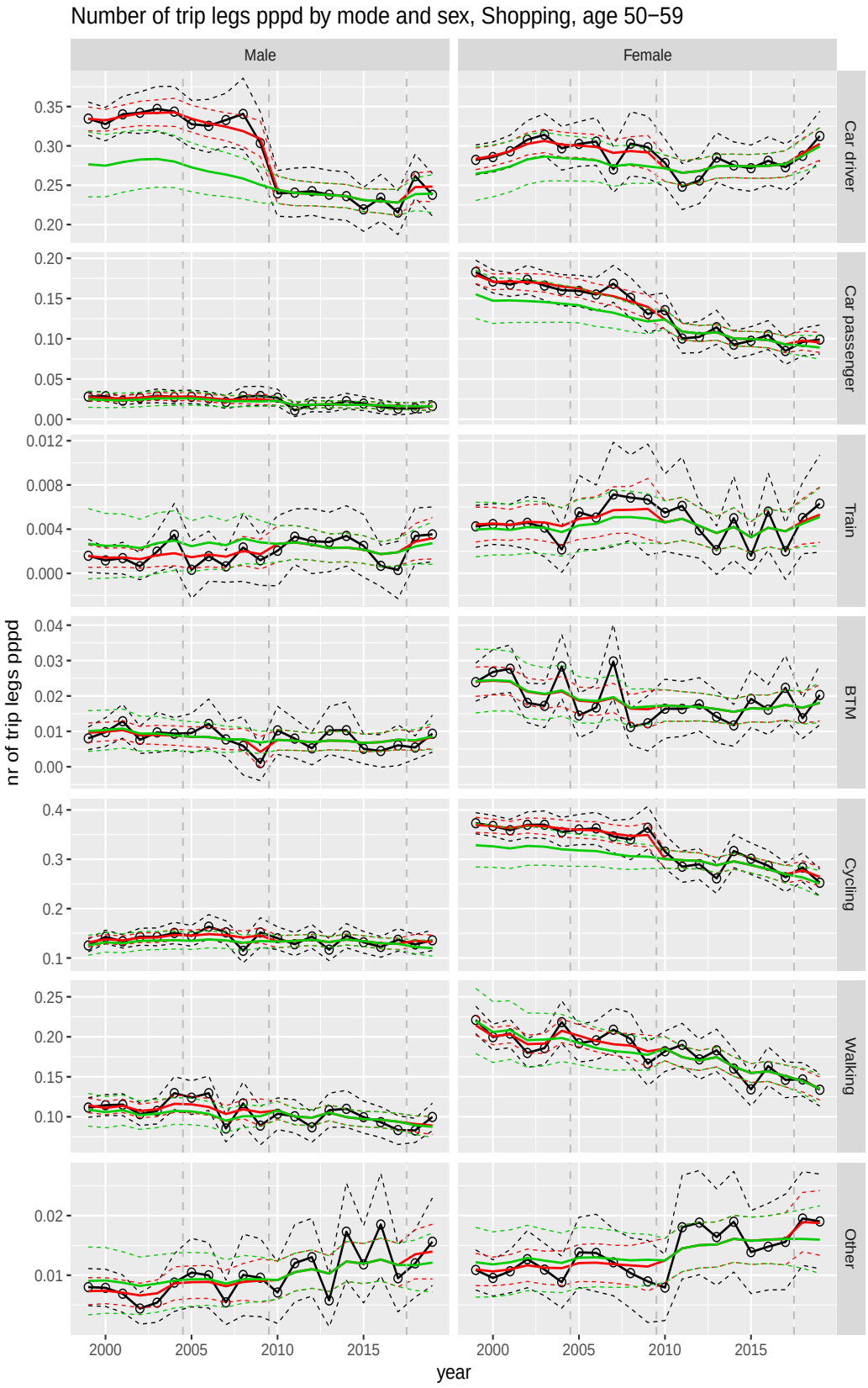

Figure A.44 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

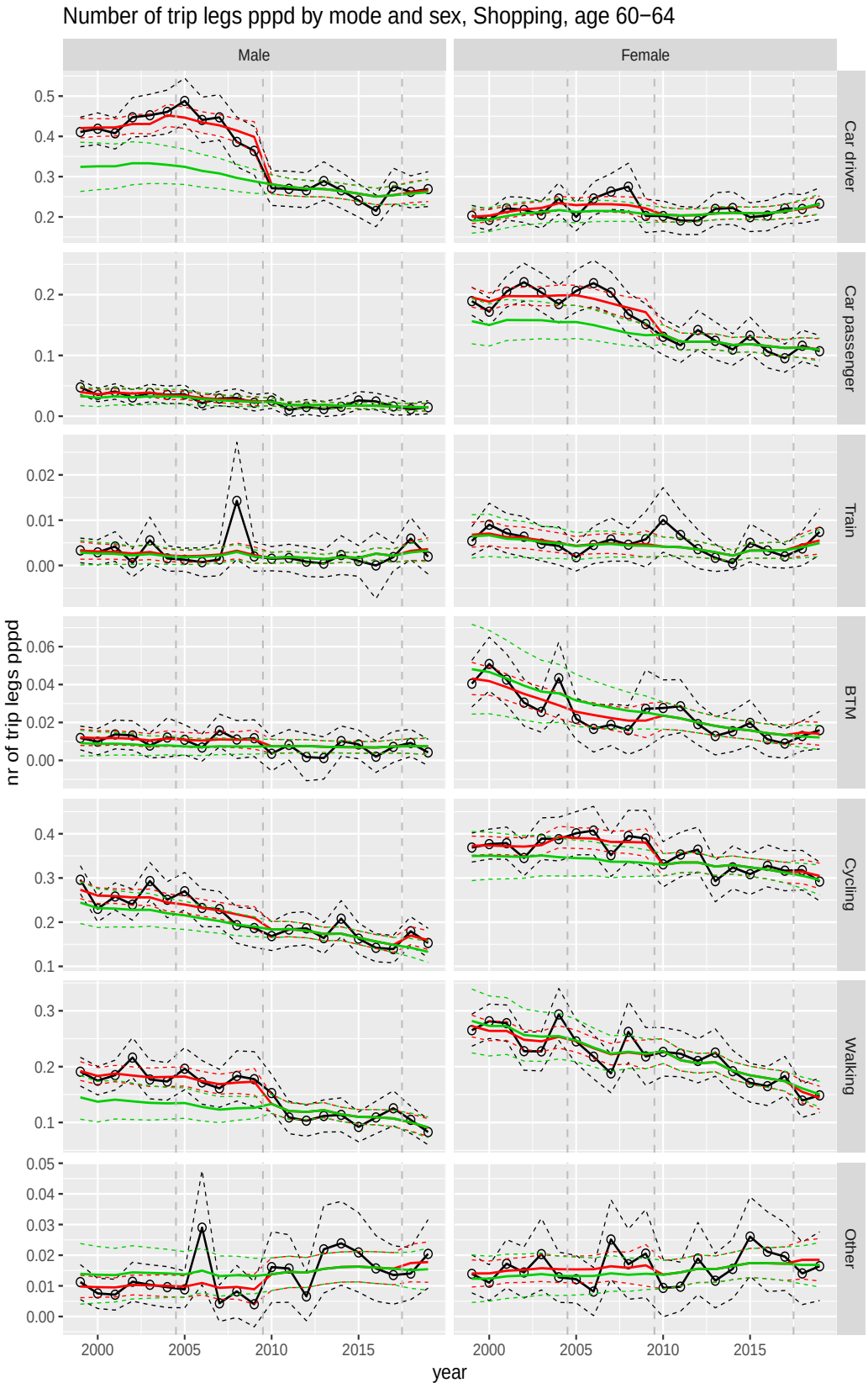

Figure A.45 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

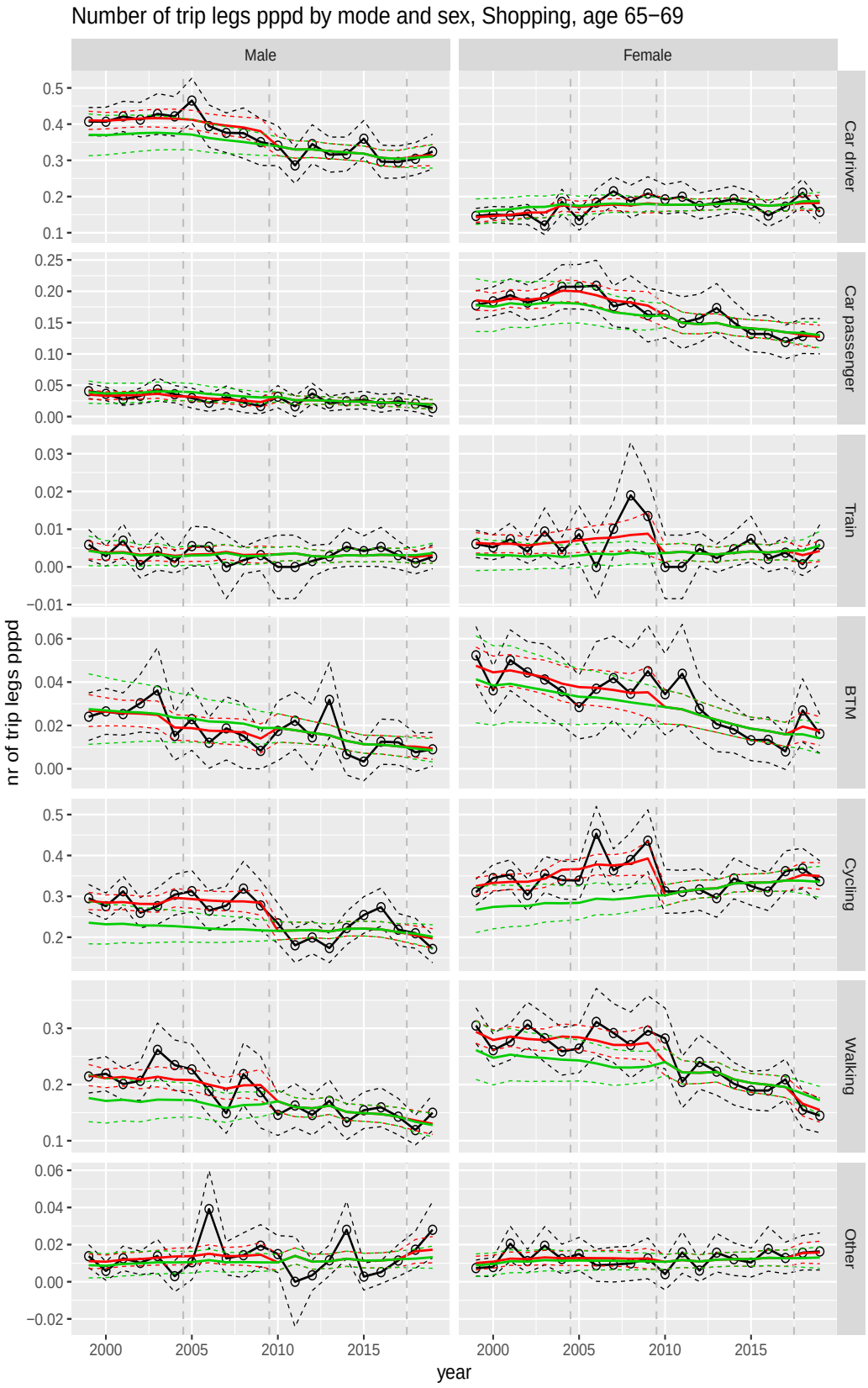

Figure A.46 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

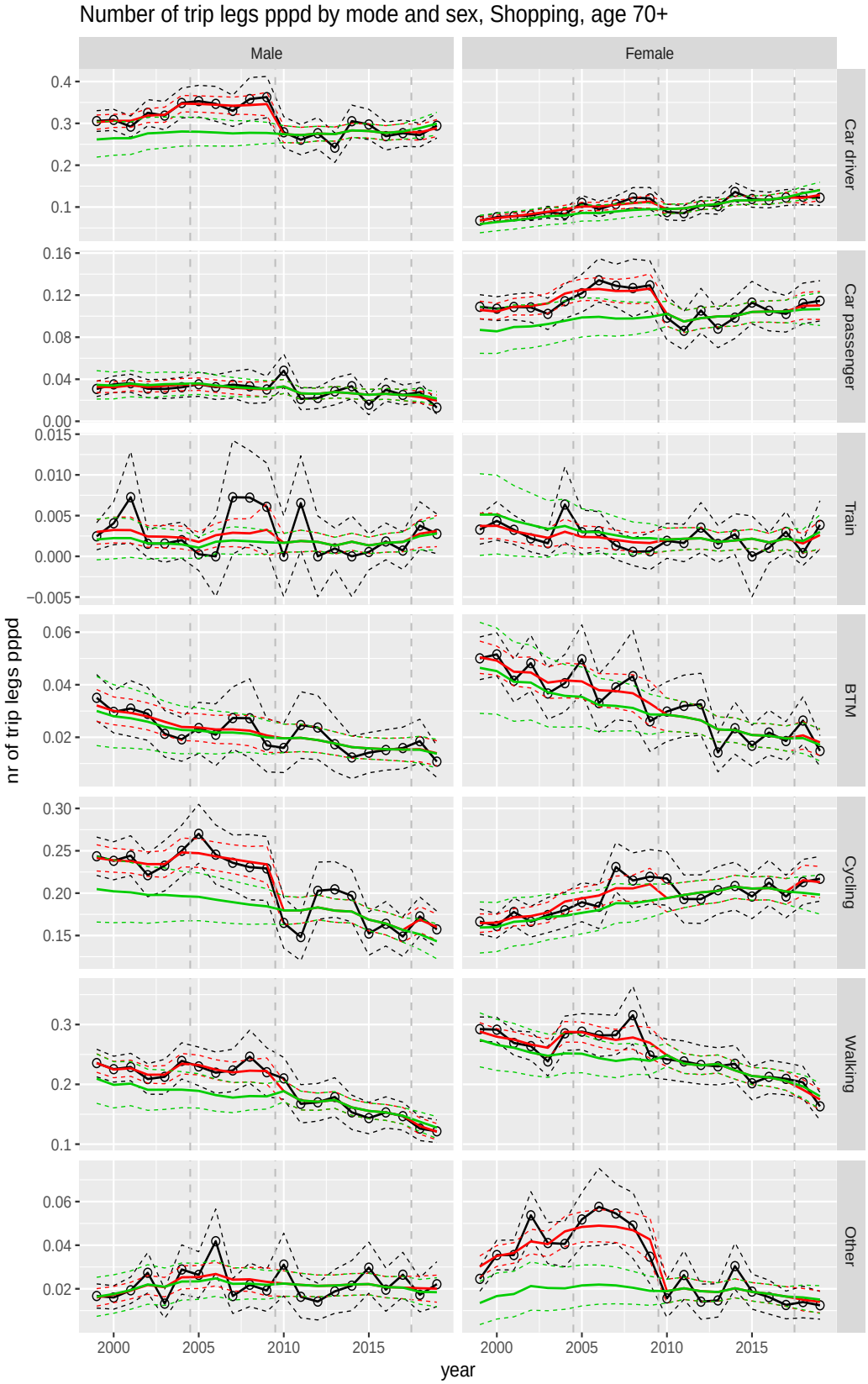

Figure A.47 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

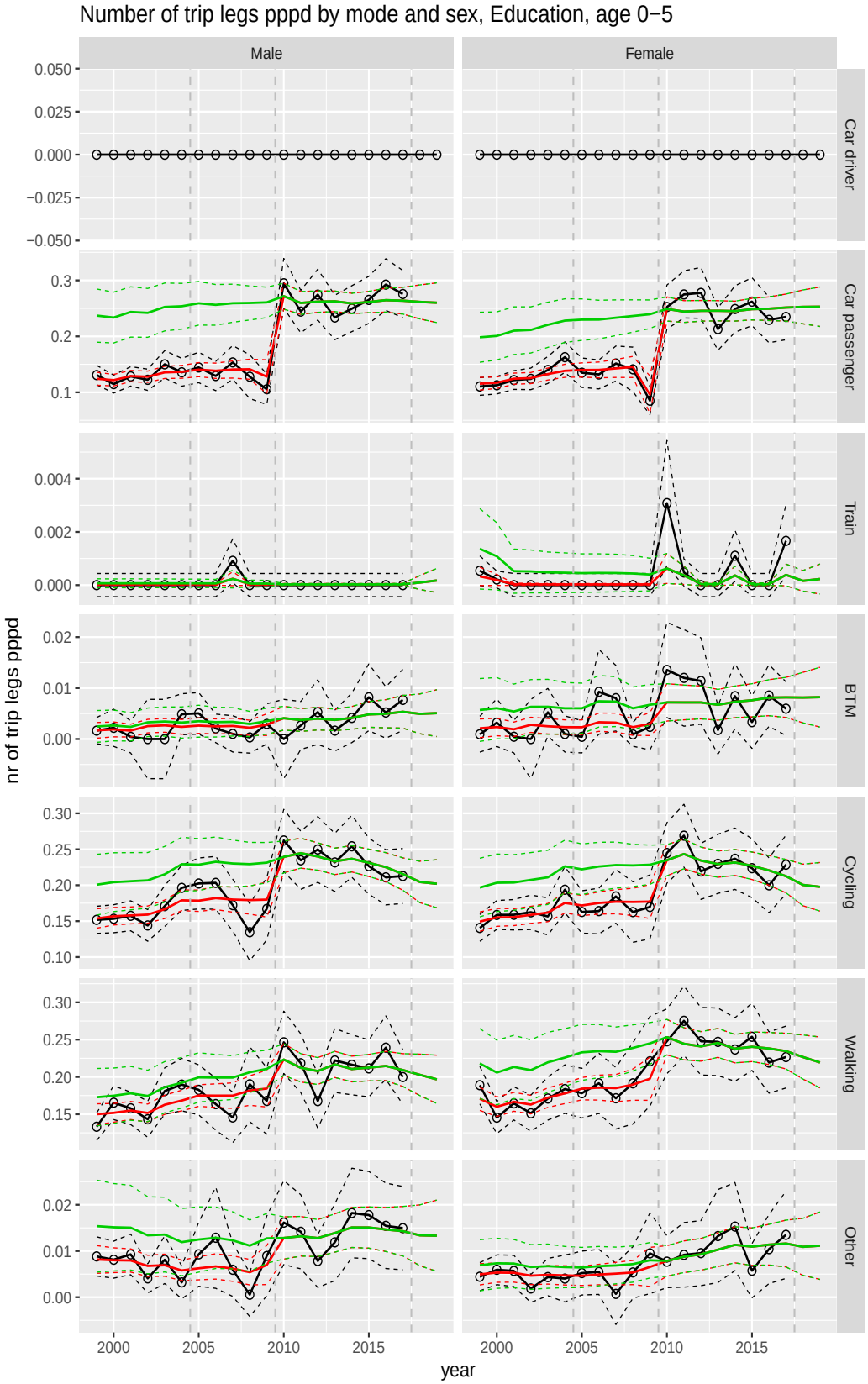

Figure A.48 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

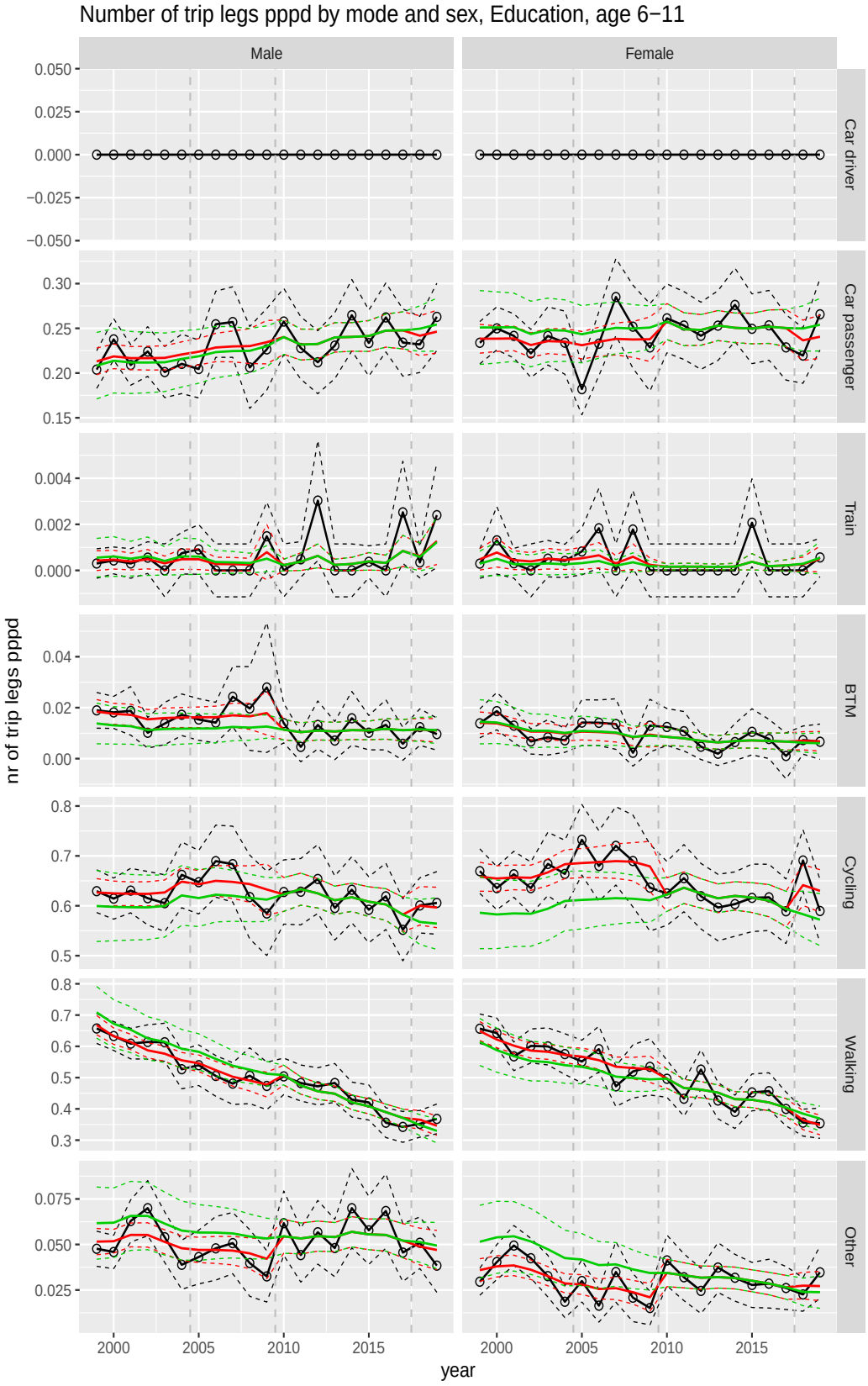

Figure A.49 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

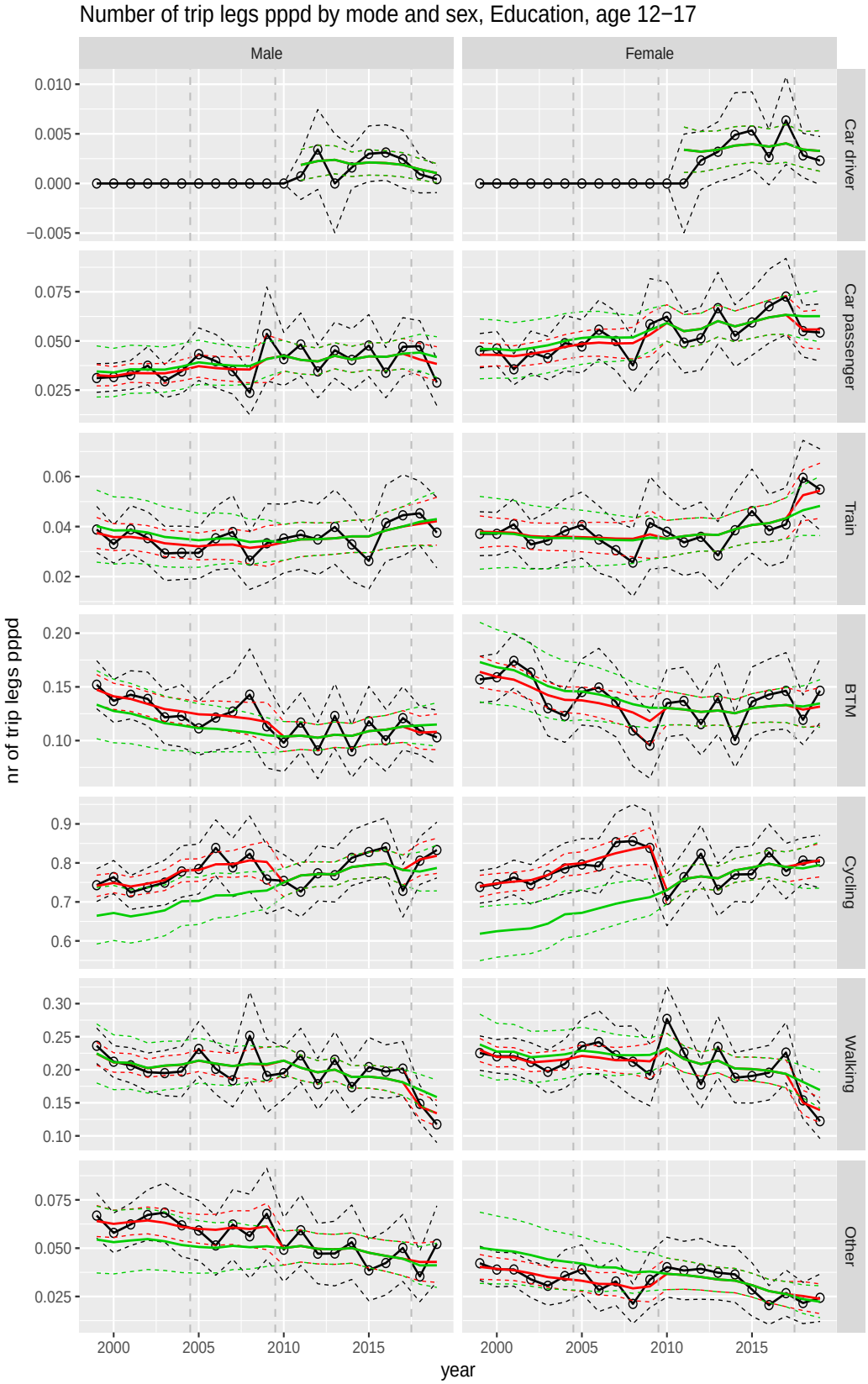

Figure A.50 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

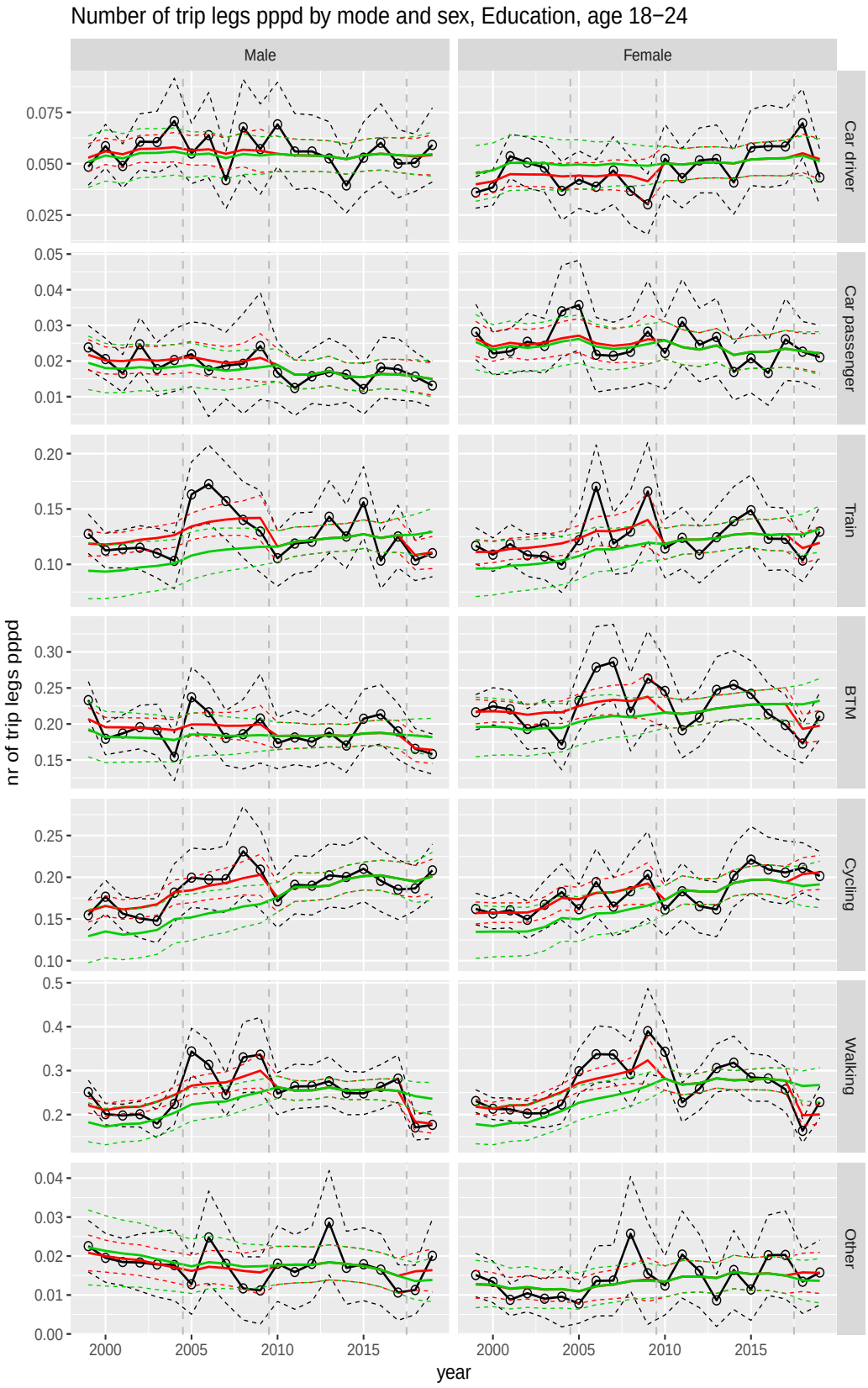

Figure A.51 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

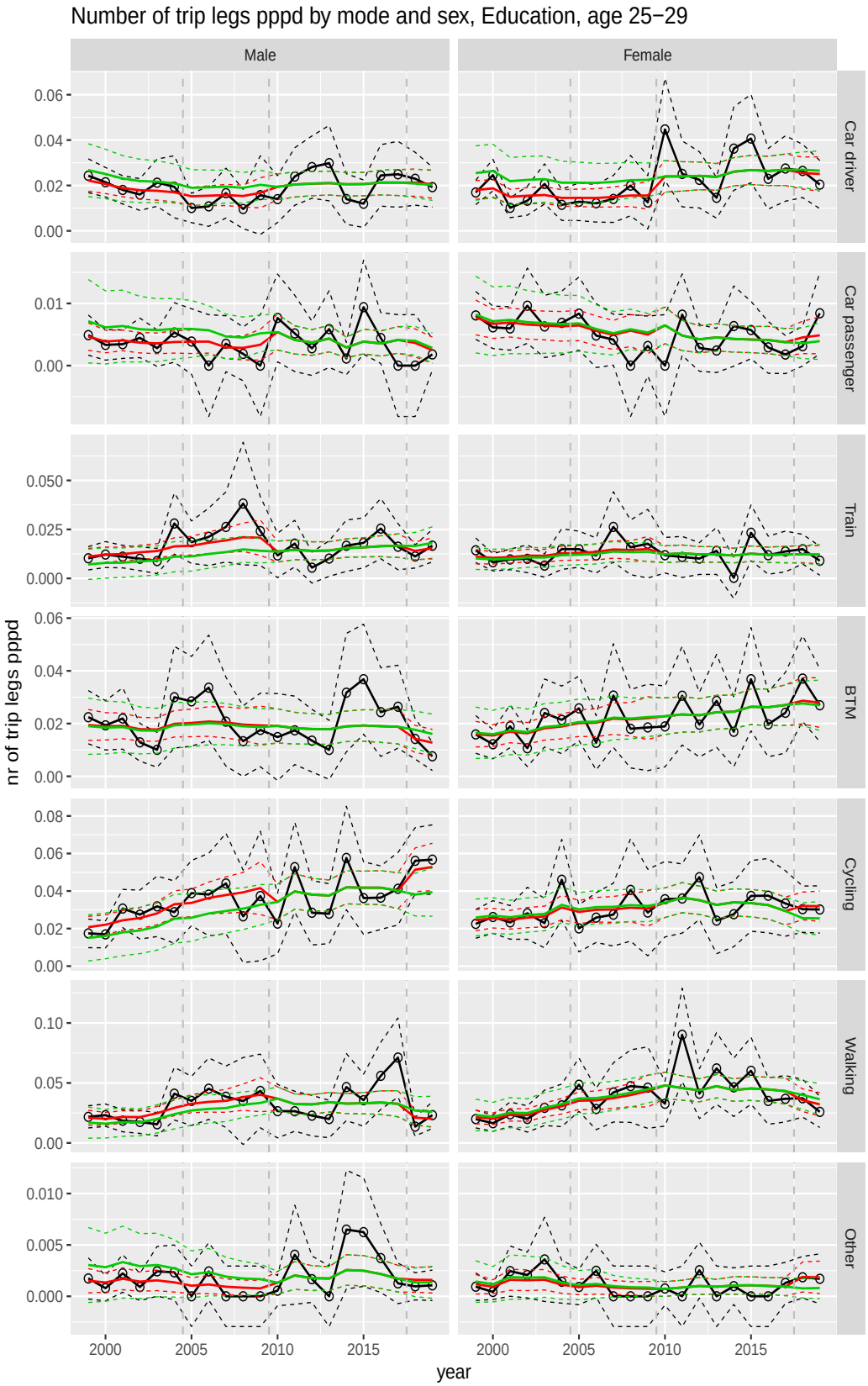

Figure A.52 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

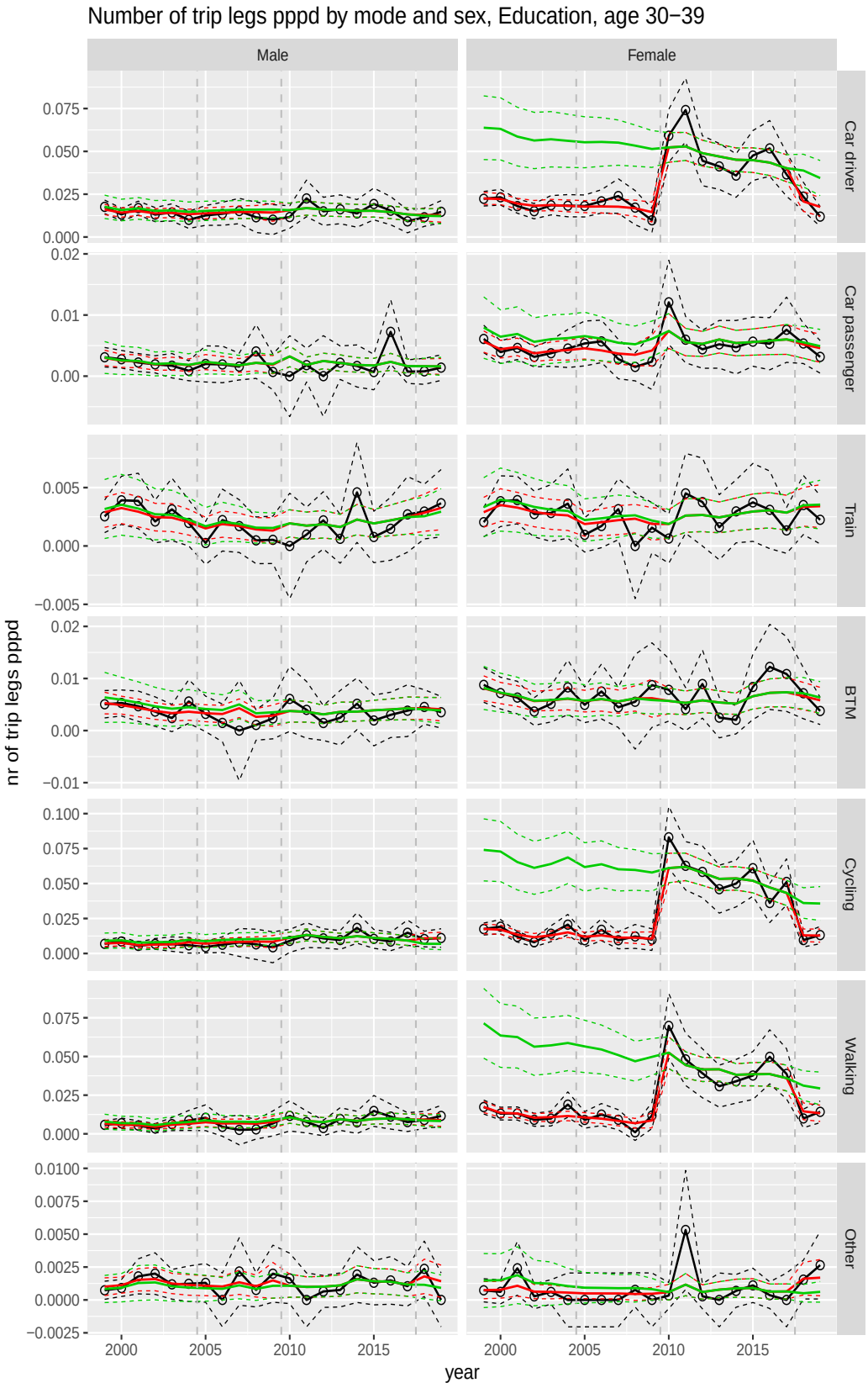

Figure A.53 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

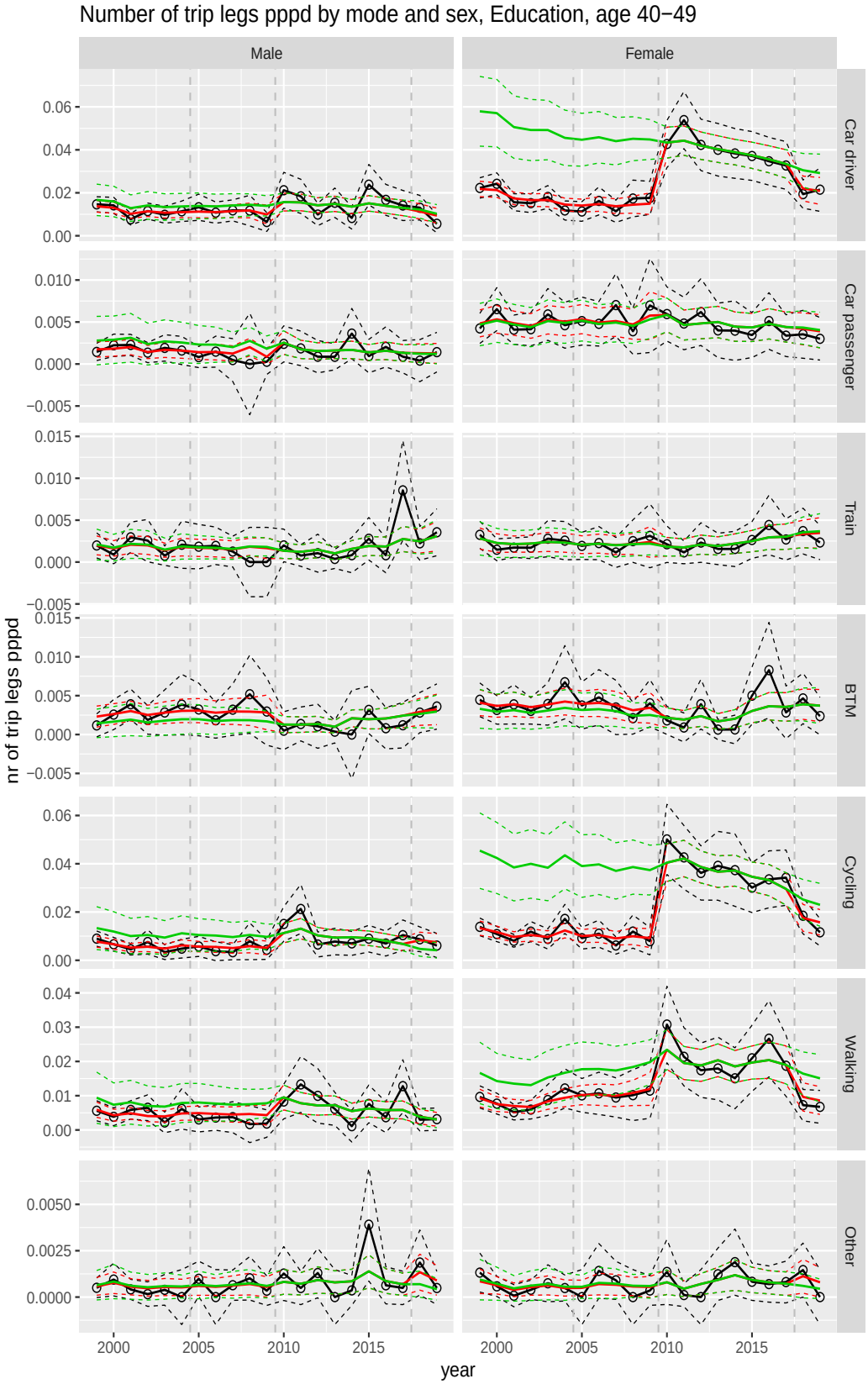

Figure A.54 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

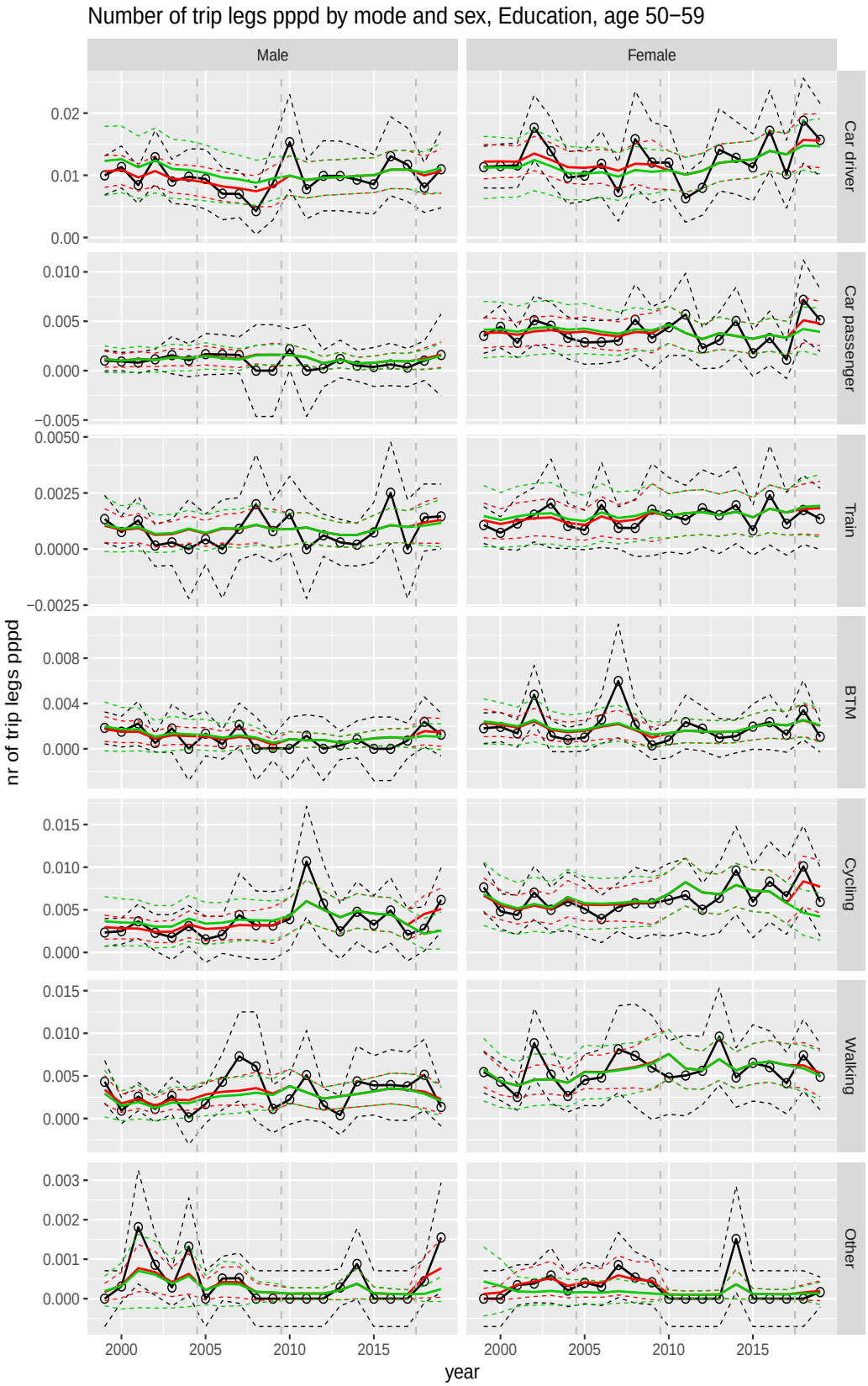

Figure A.55 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

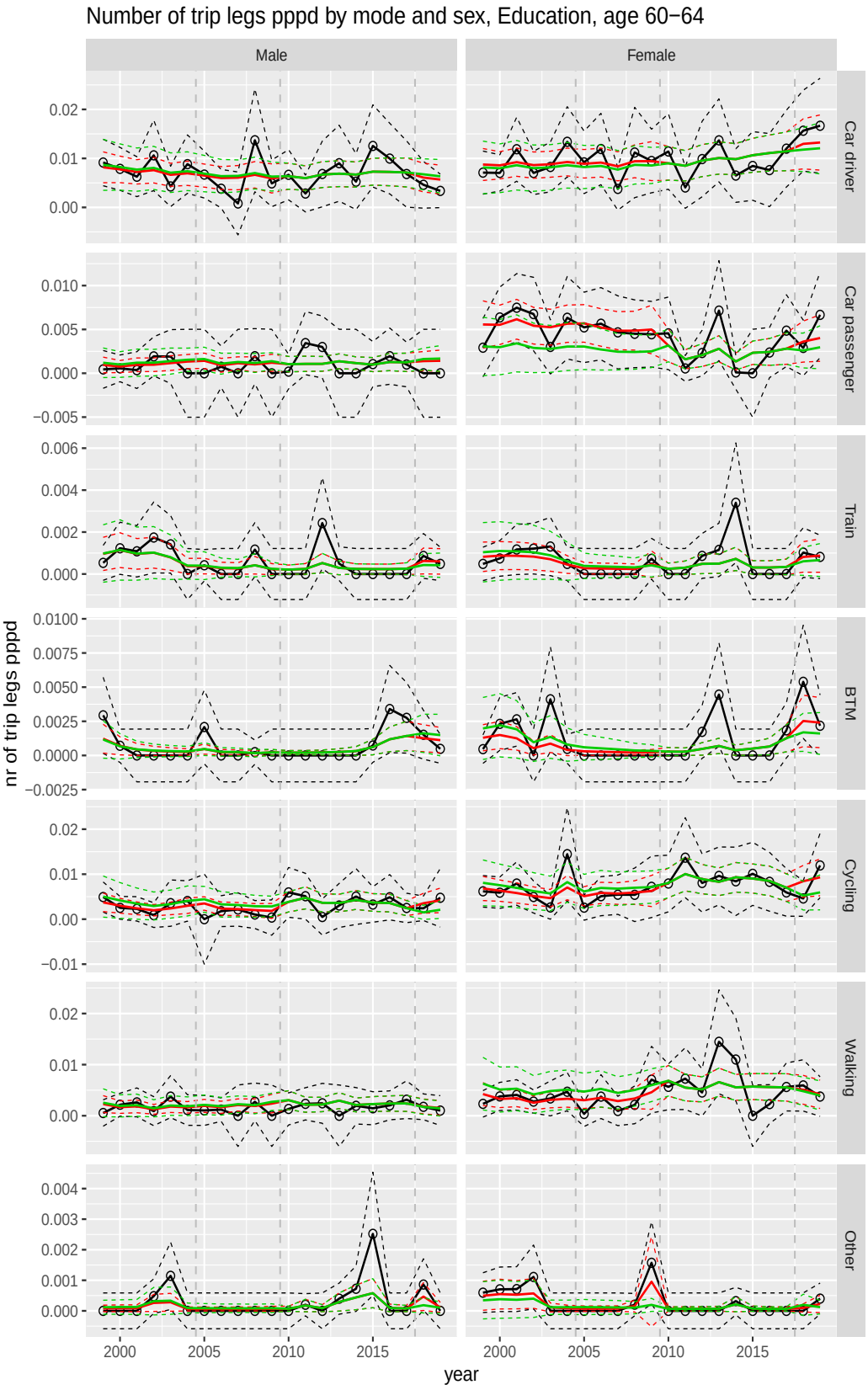

Figure A.56 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

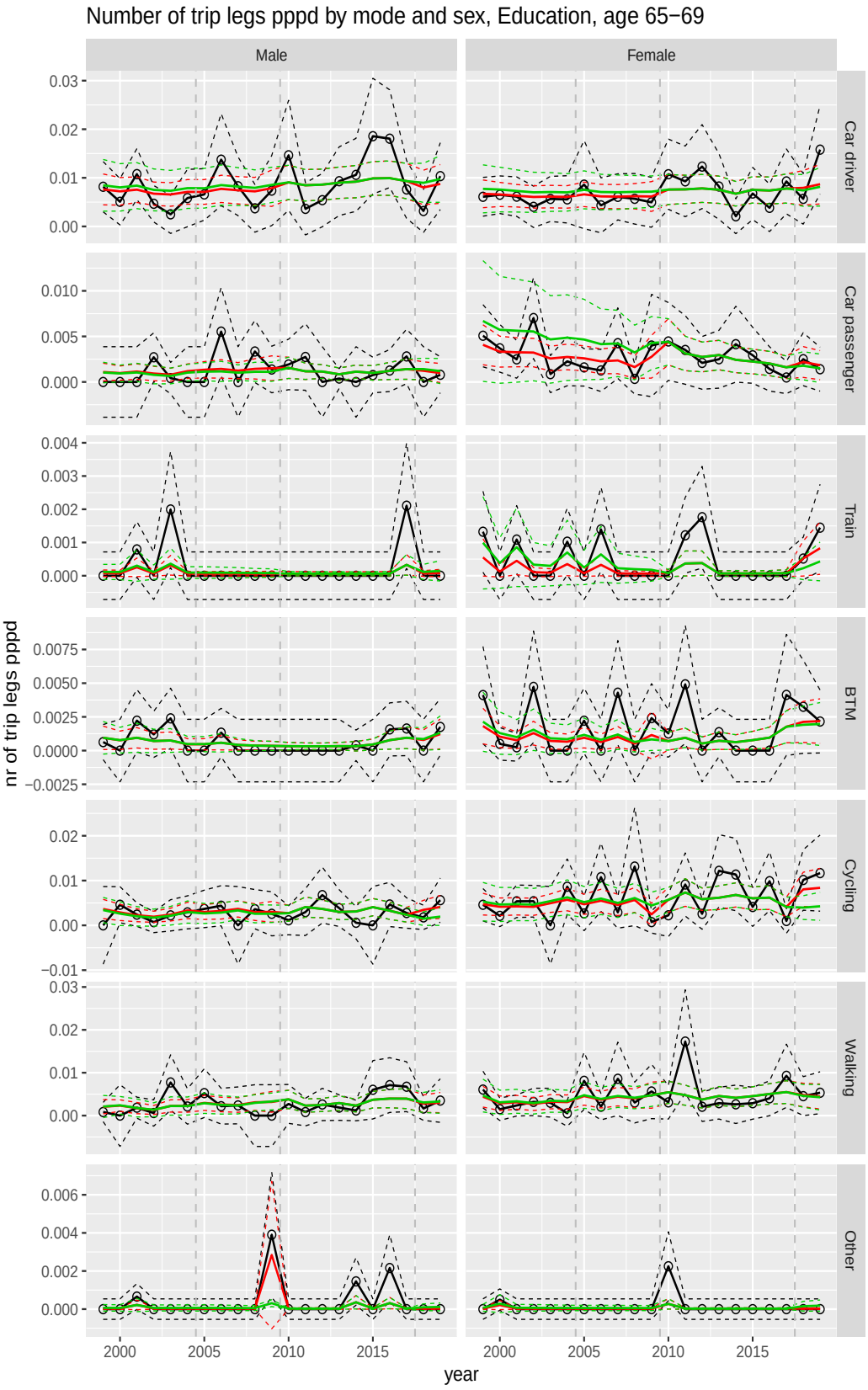

Figure A.57 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

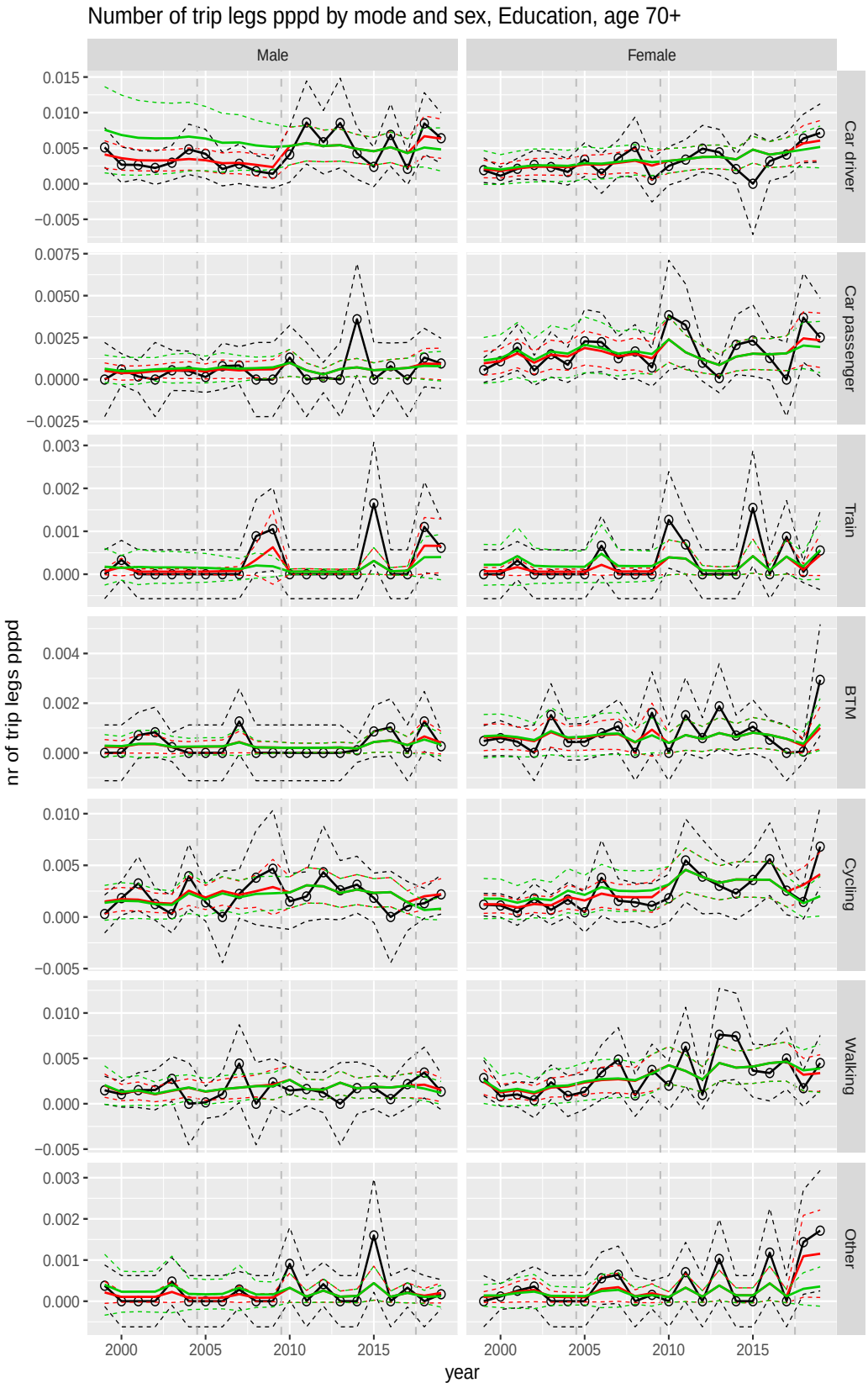

Figure A.58 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

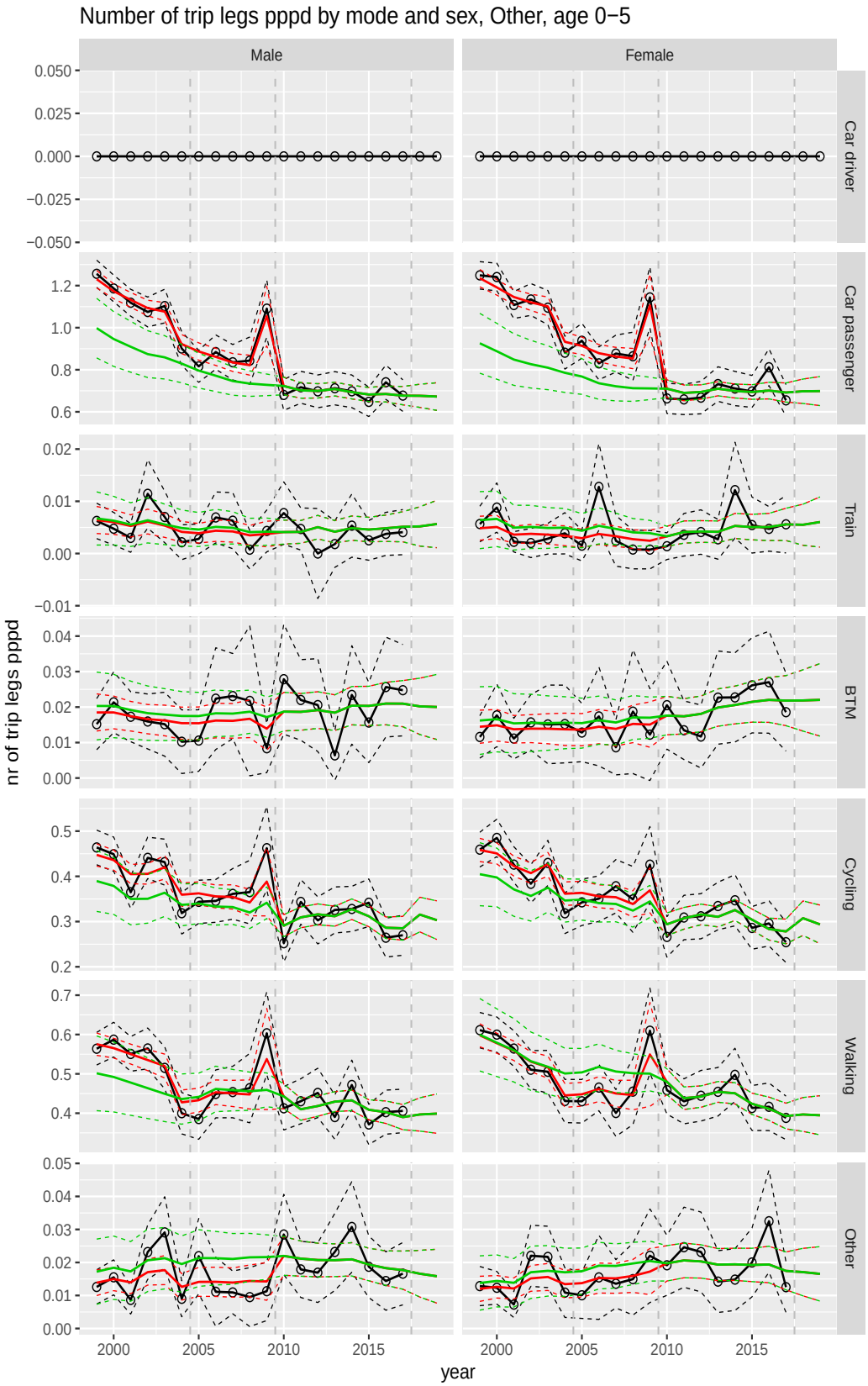

Figure A.59 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

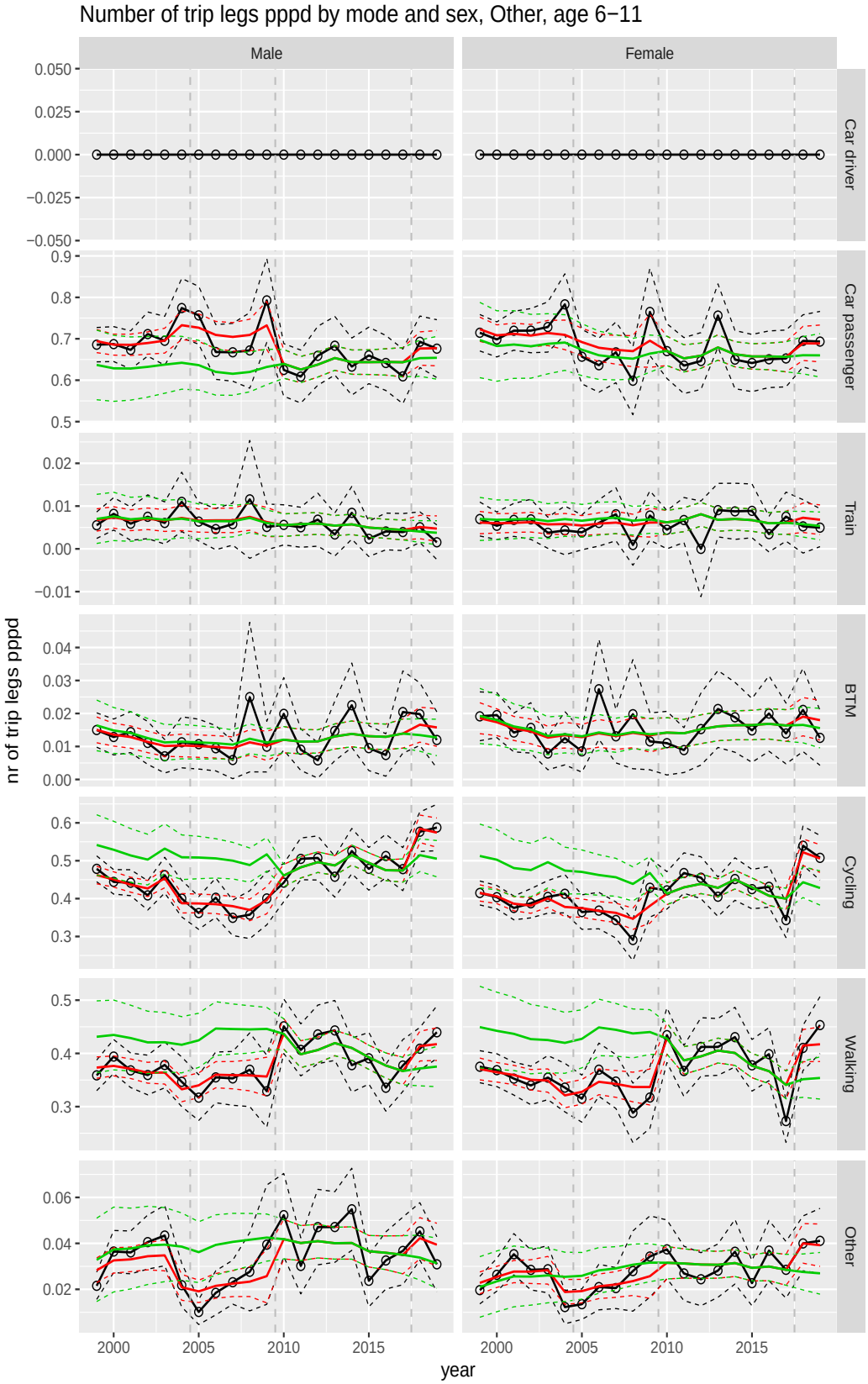

Figure A.60 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure A.61 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

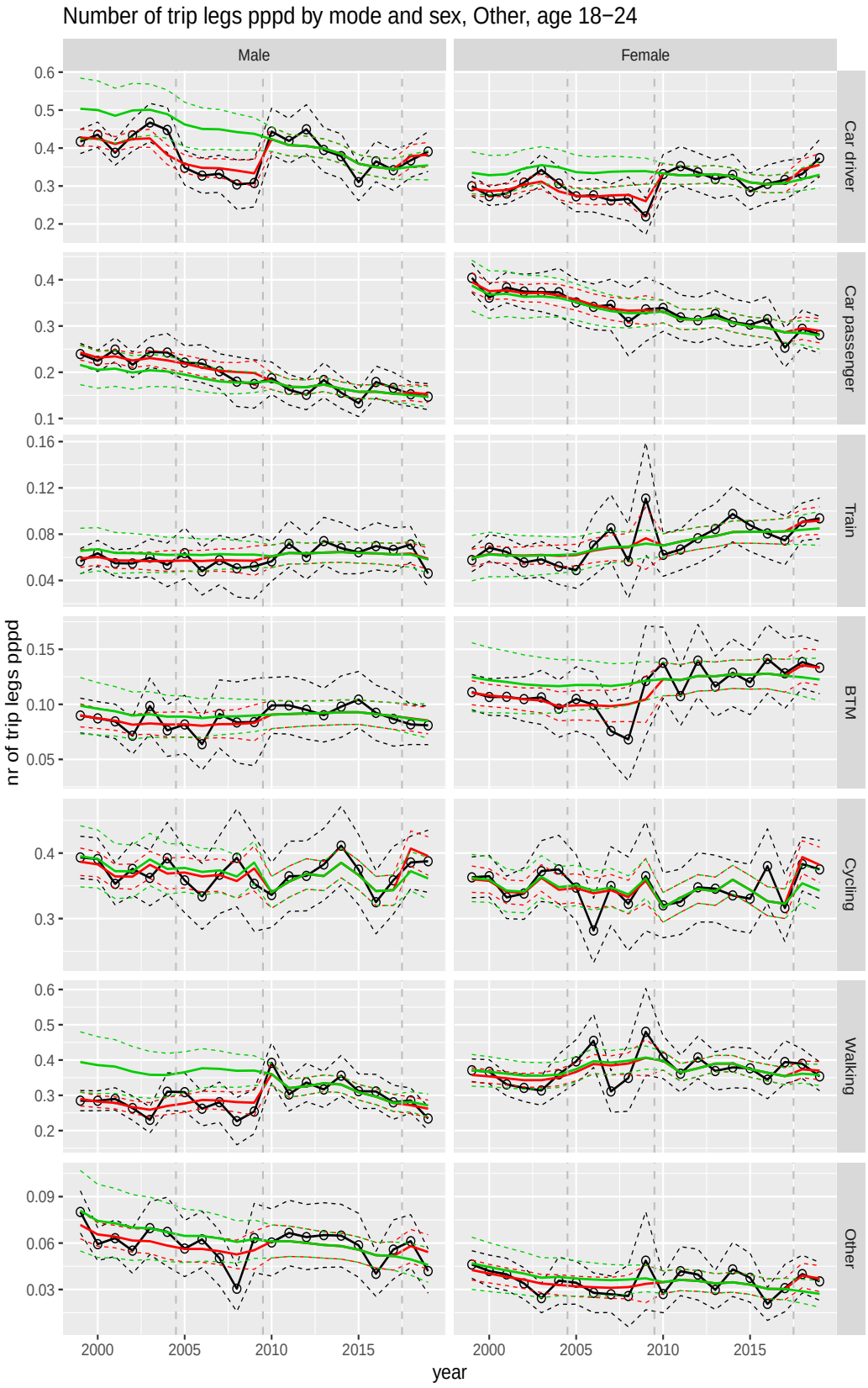

Figure A.62 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

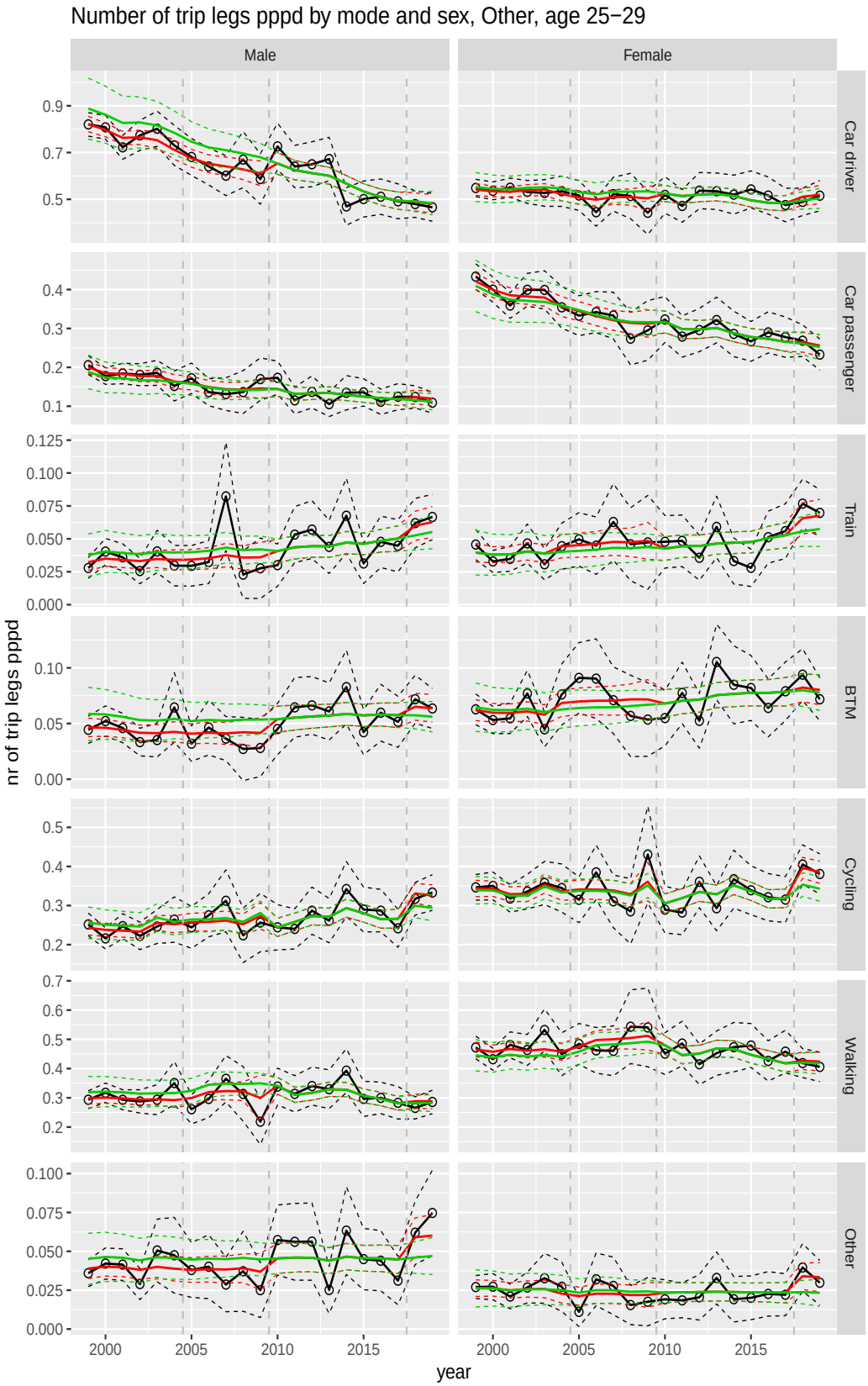

Figure A.63 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

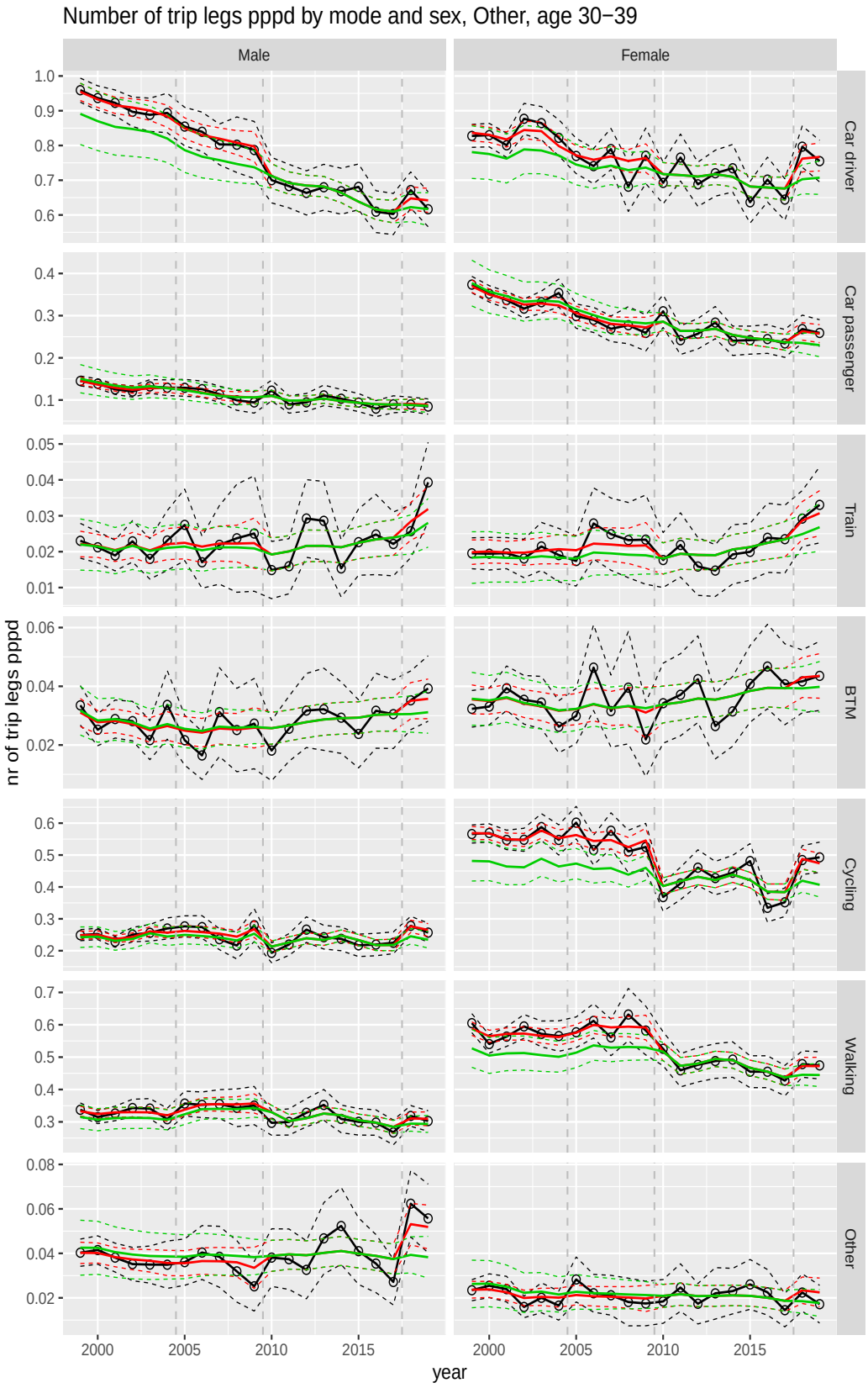

Figure A.64 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

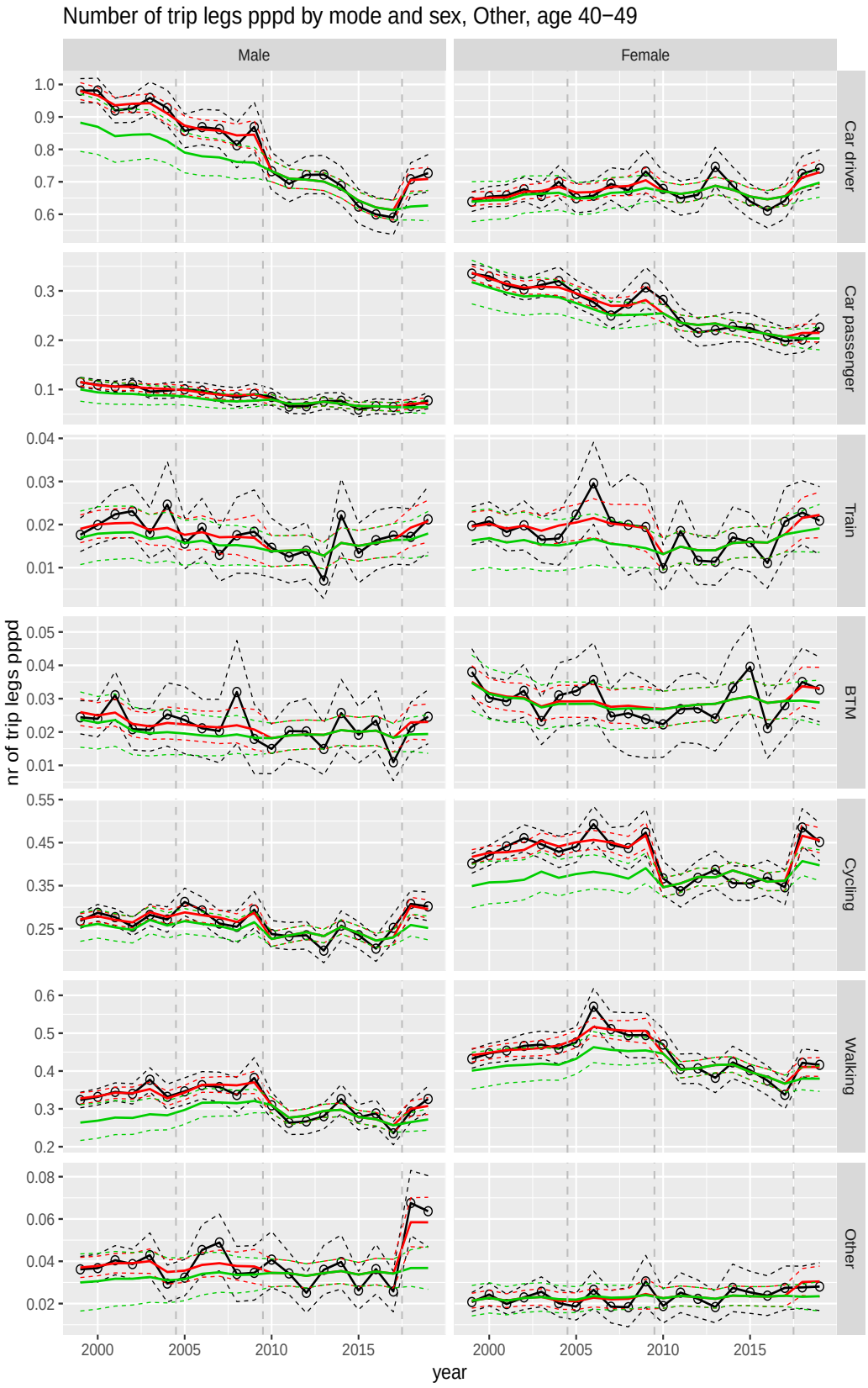

Figure A.65 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

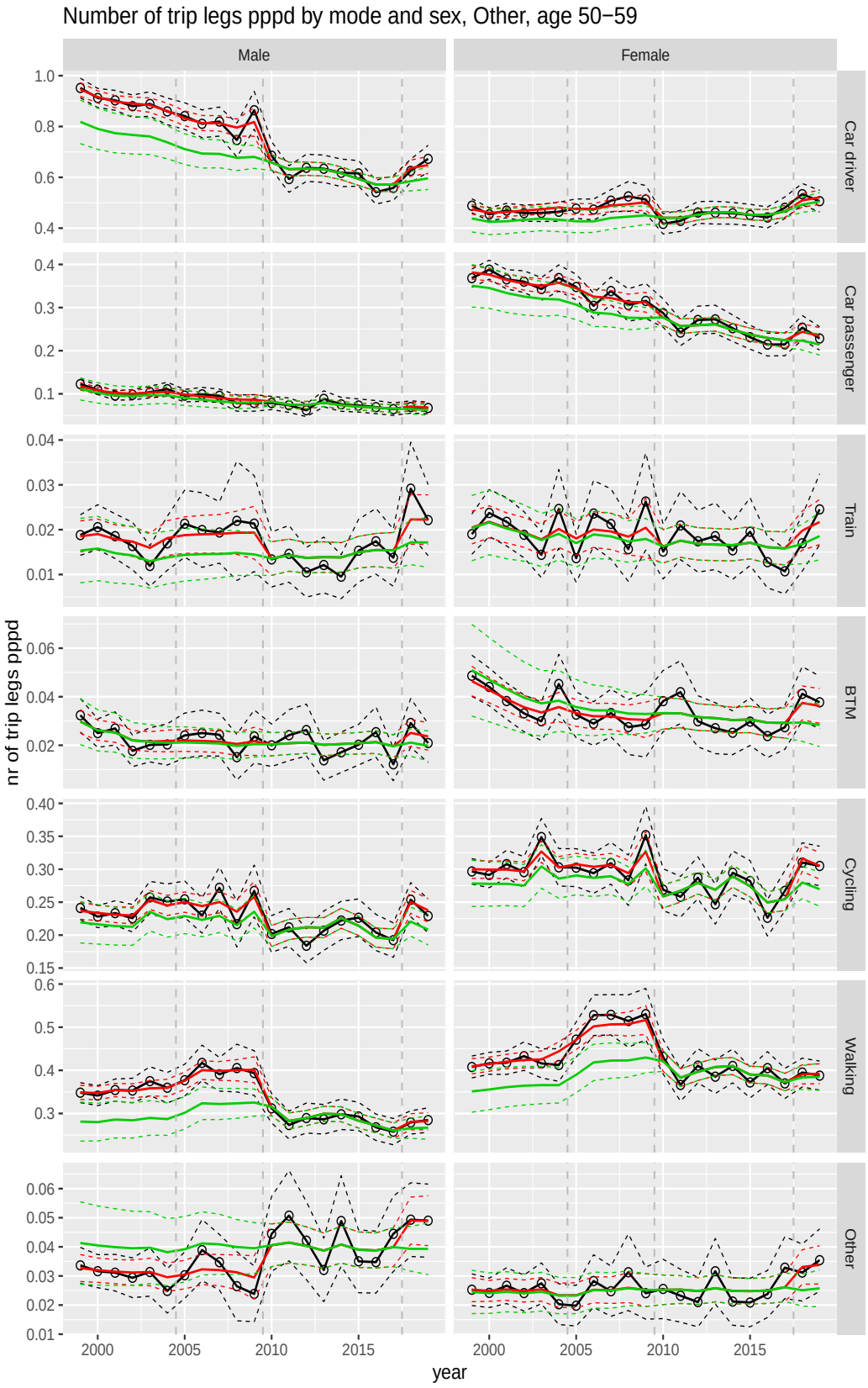

Figure A.66 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

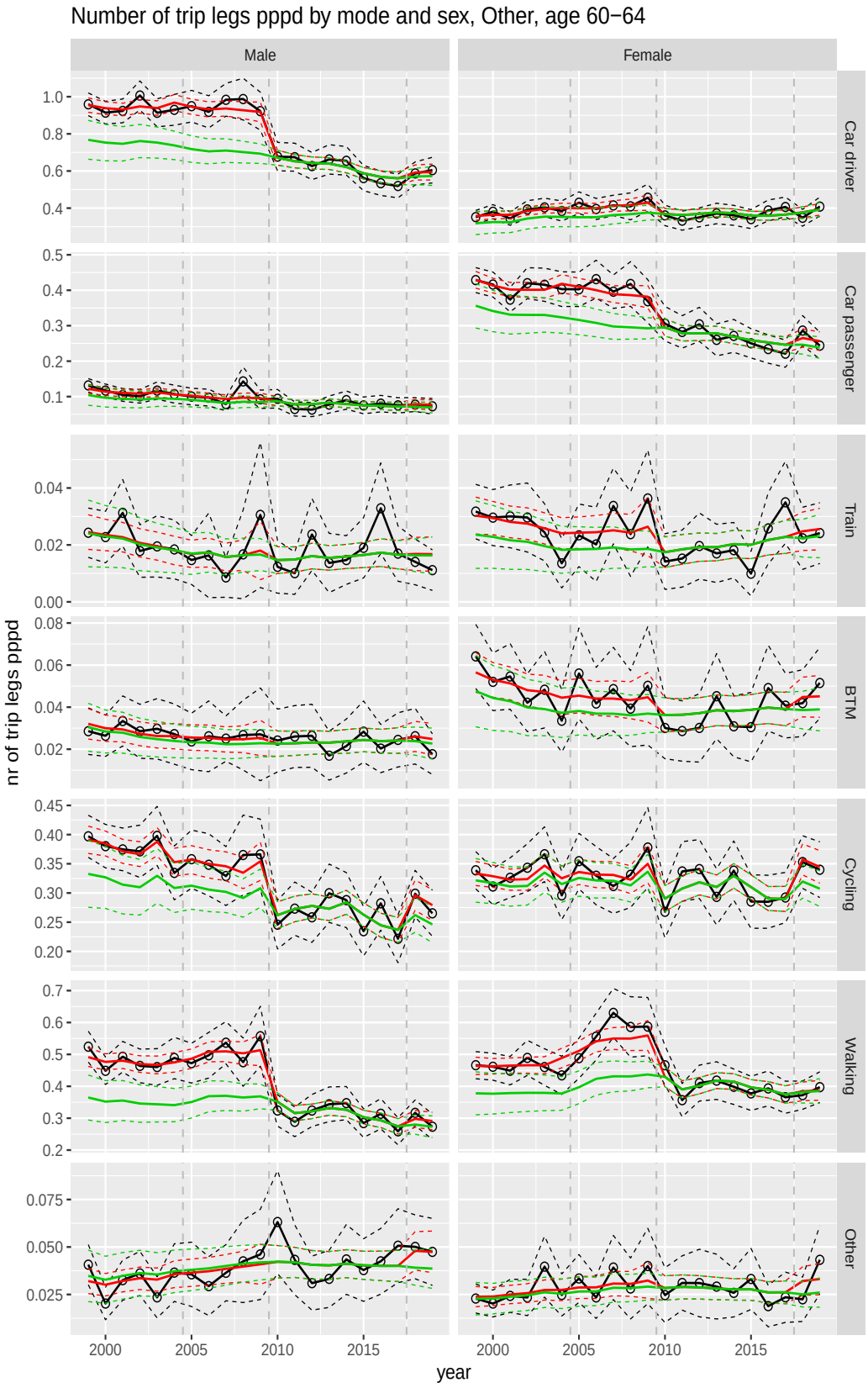

Figure A.67 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

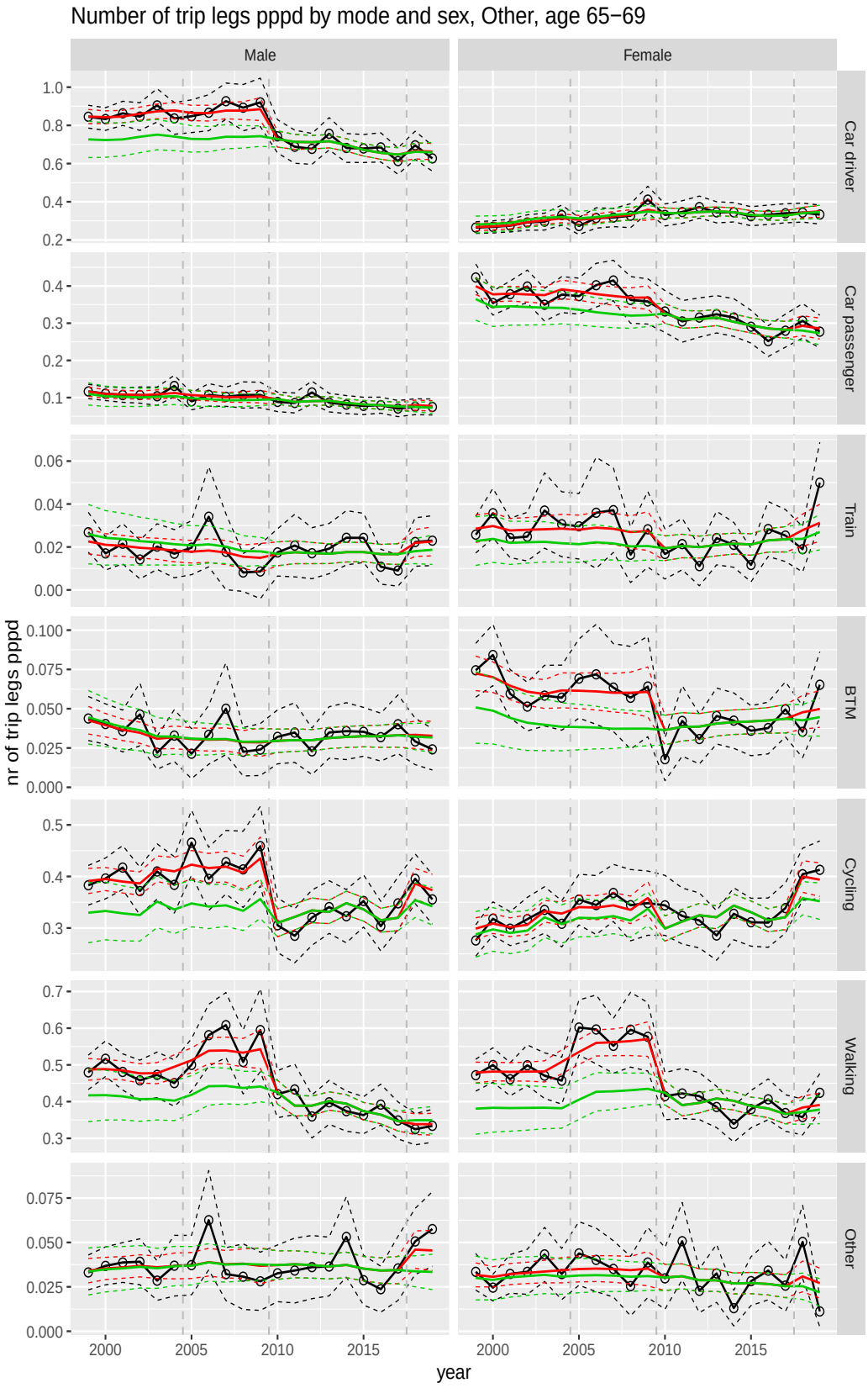

Figure A.68 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

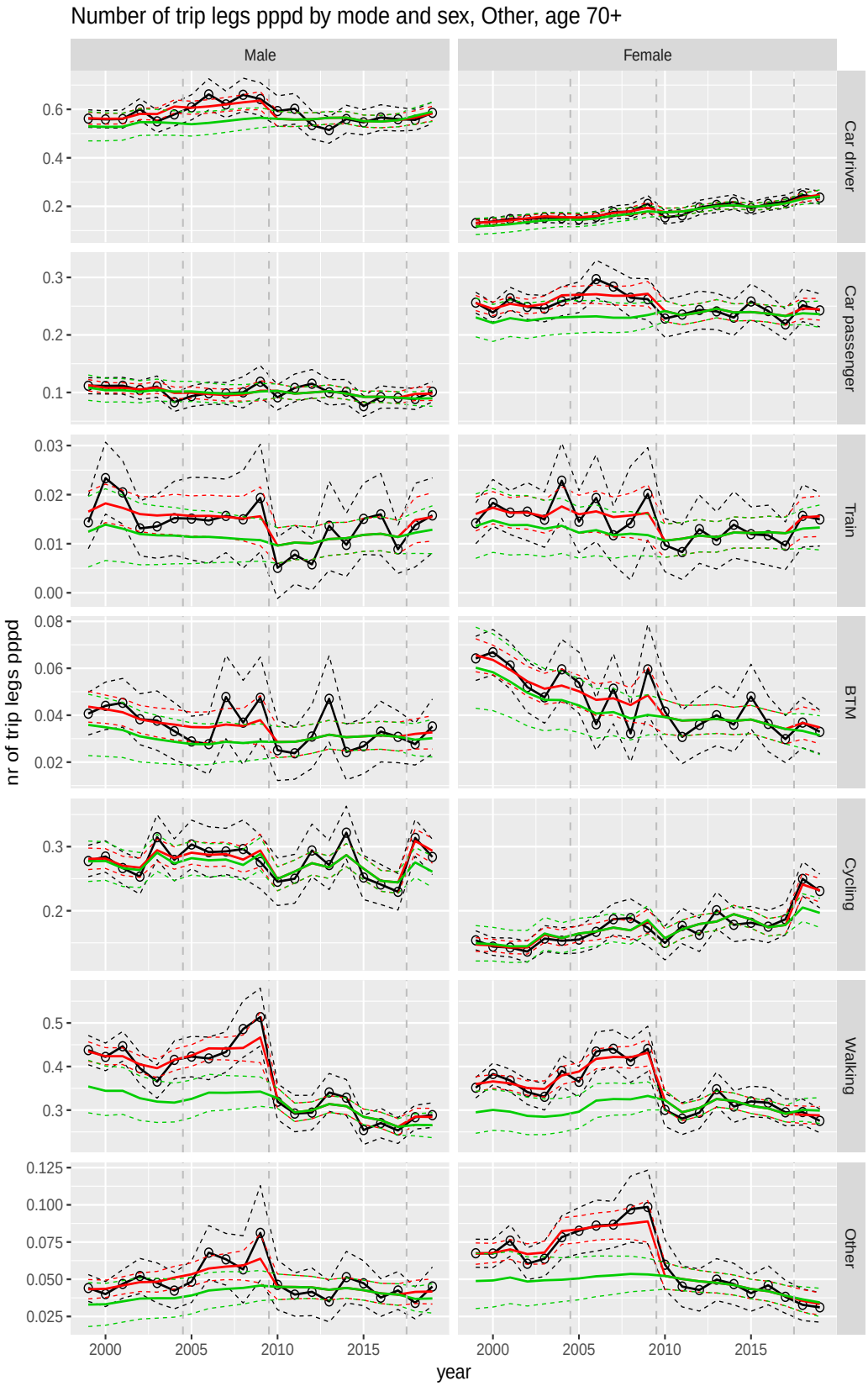

Figure A.69 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

A.2 Average distance per trip leg

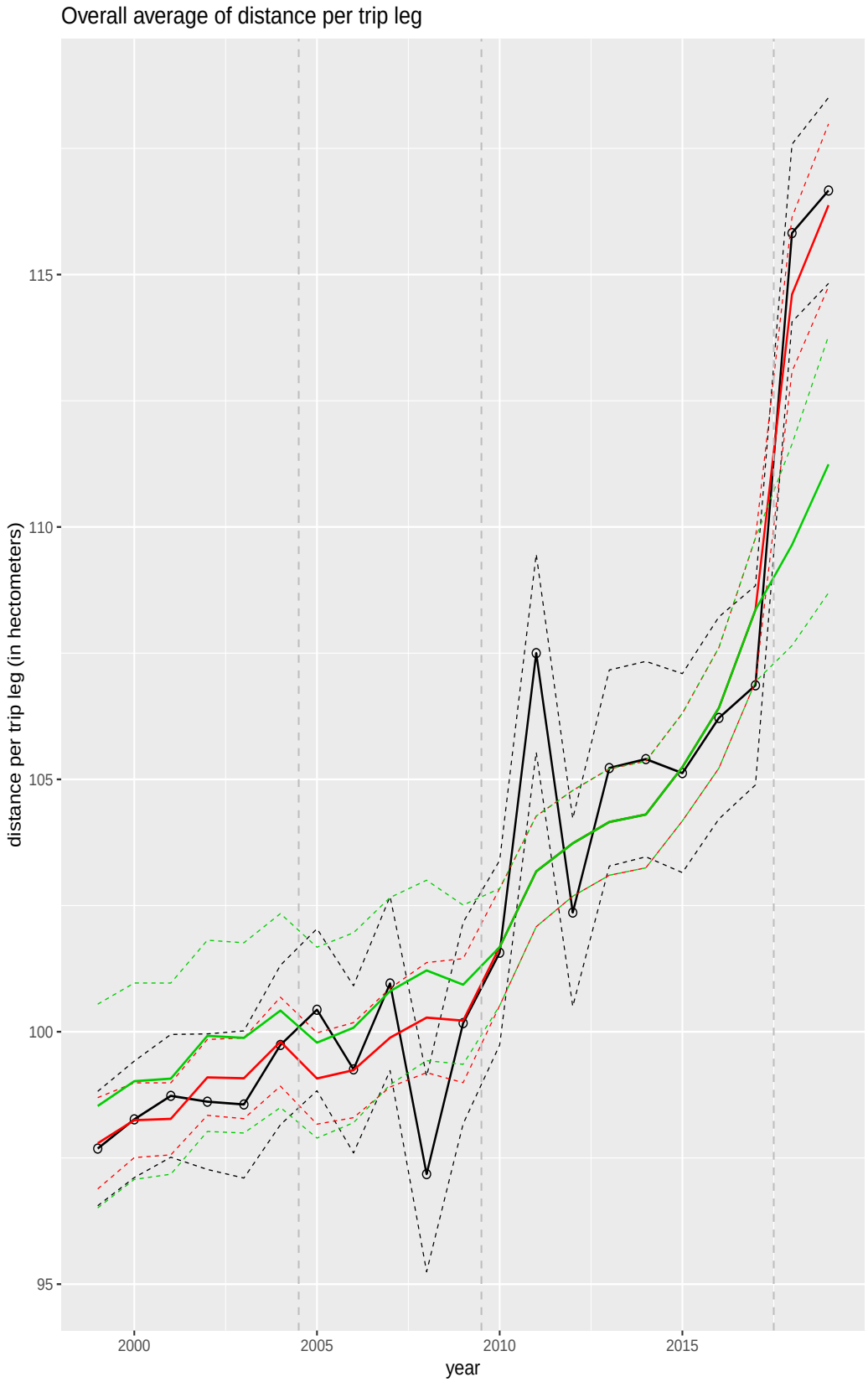

Figure A.70 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

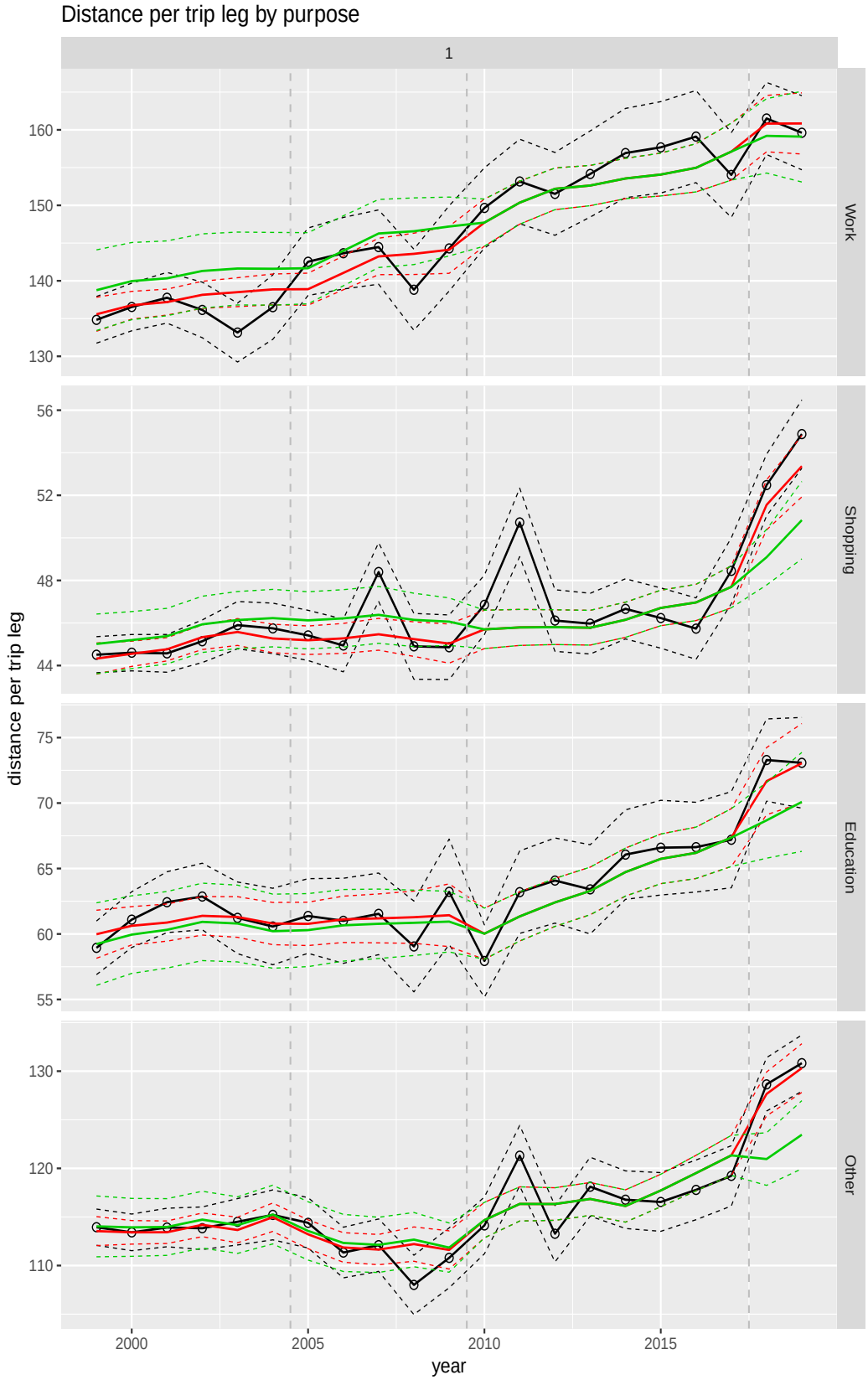

Figure A.71 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

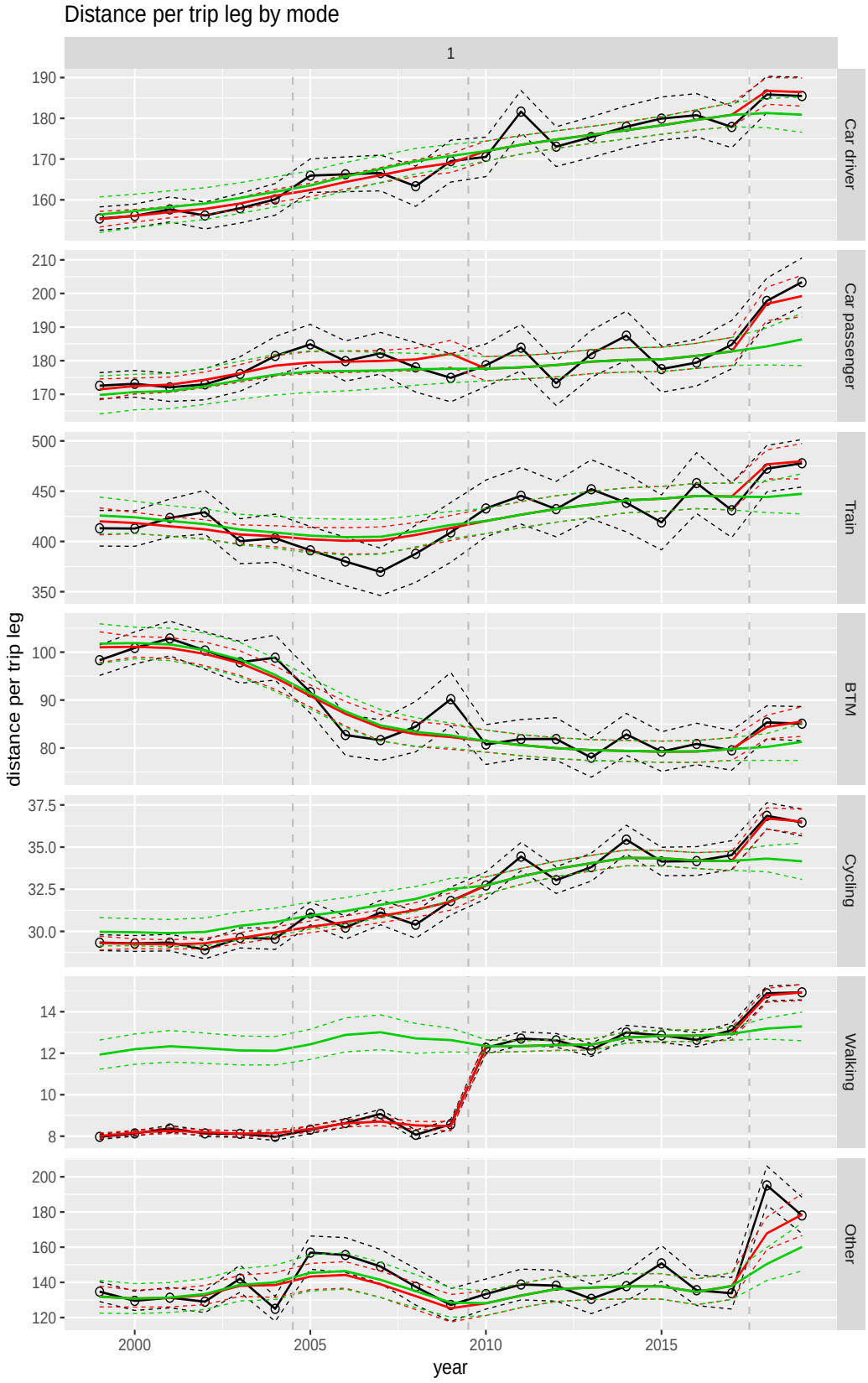

Figure A.72 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure A.73 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

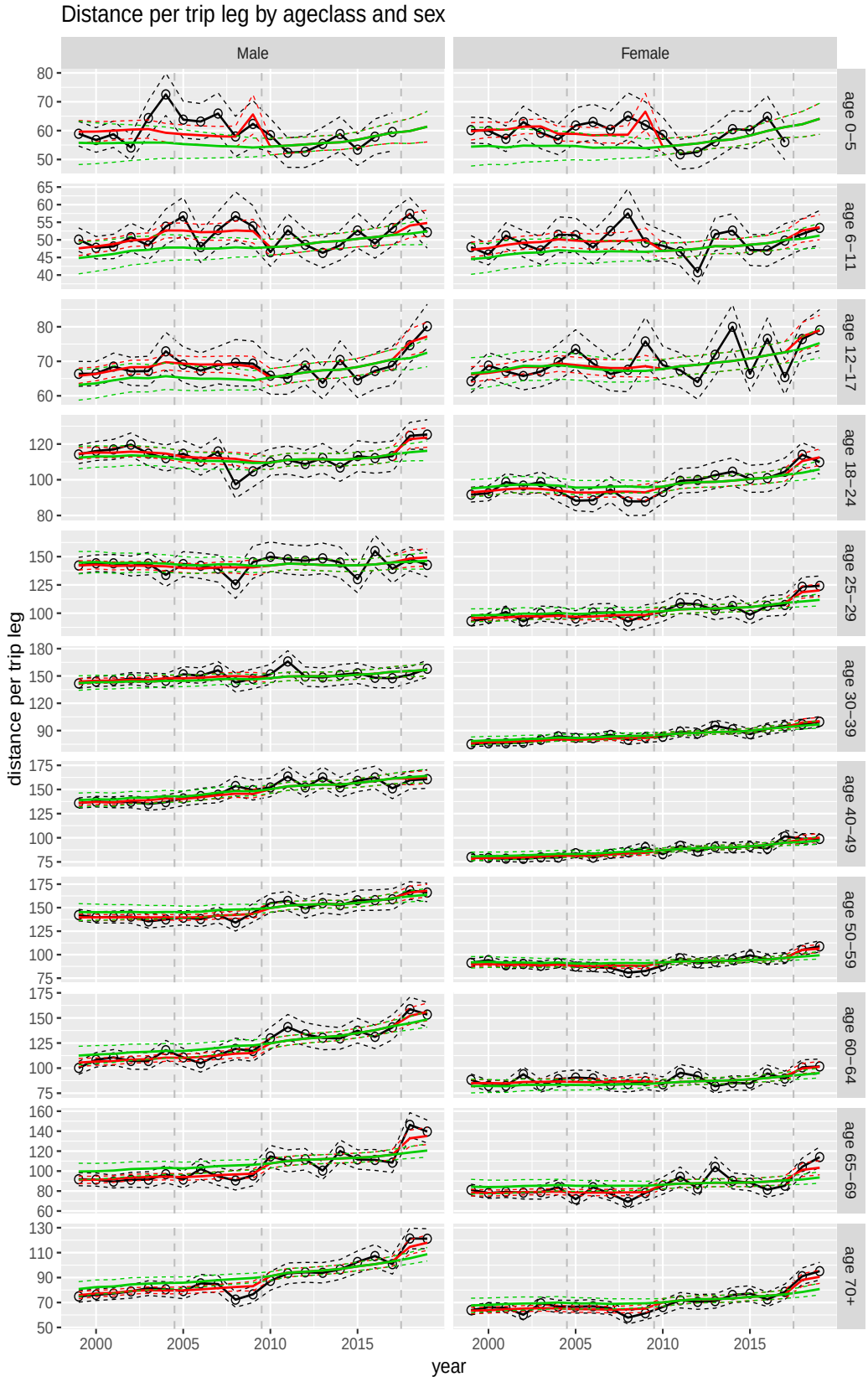

Figure A.74 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

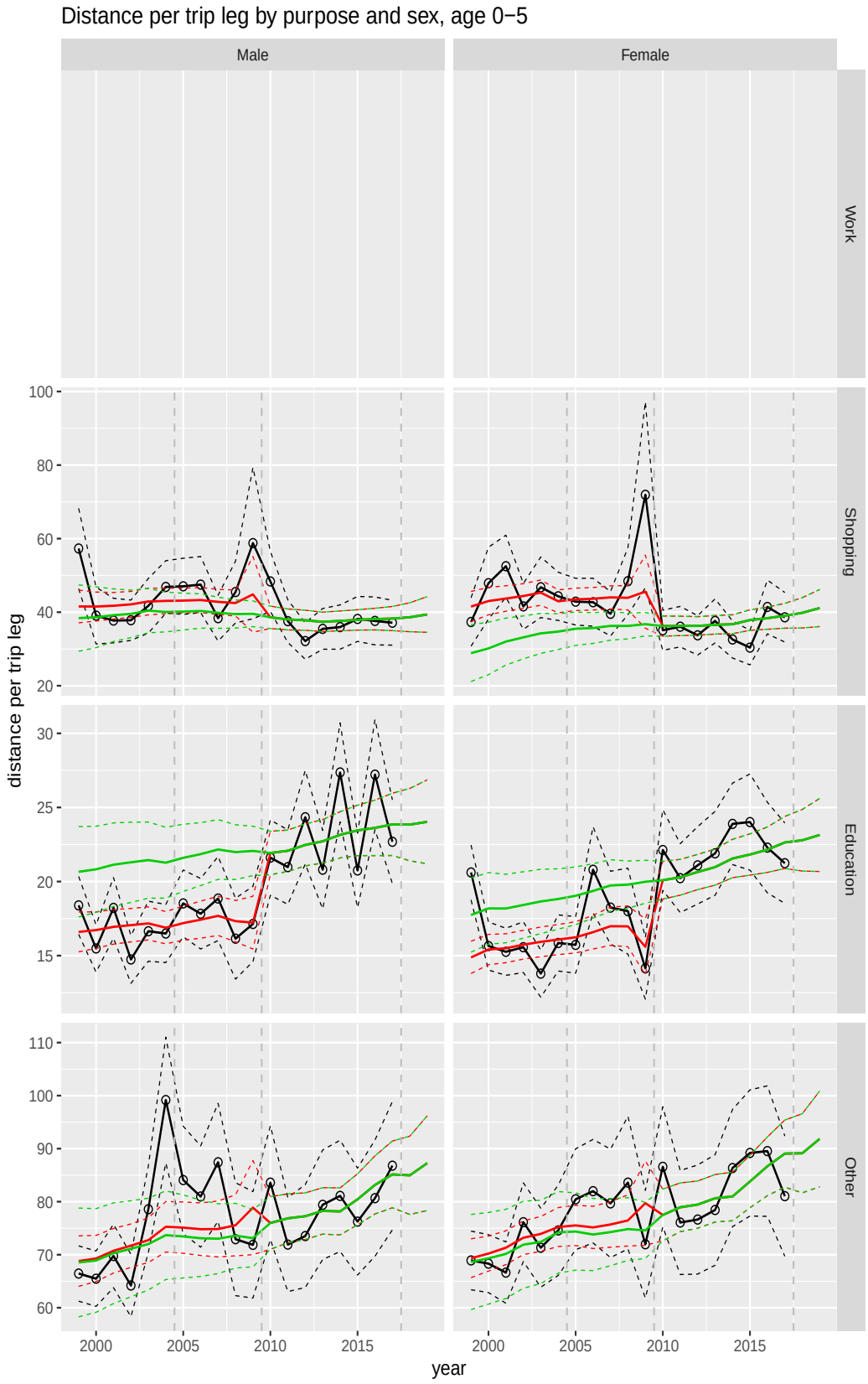

Figure A.75 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

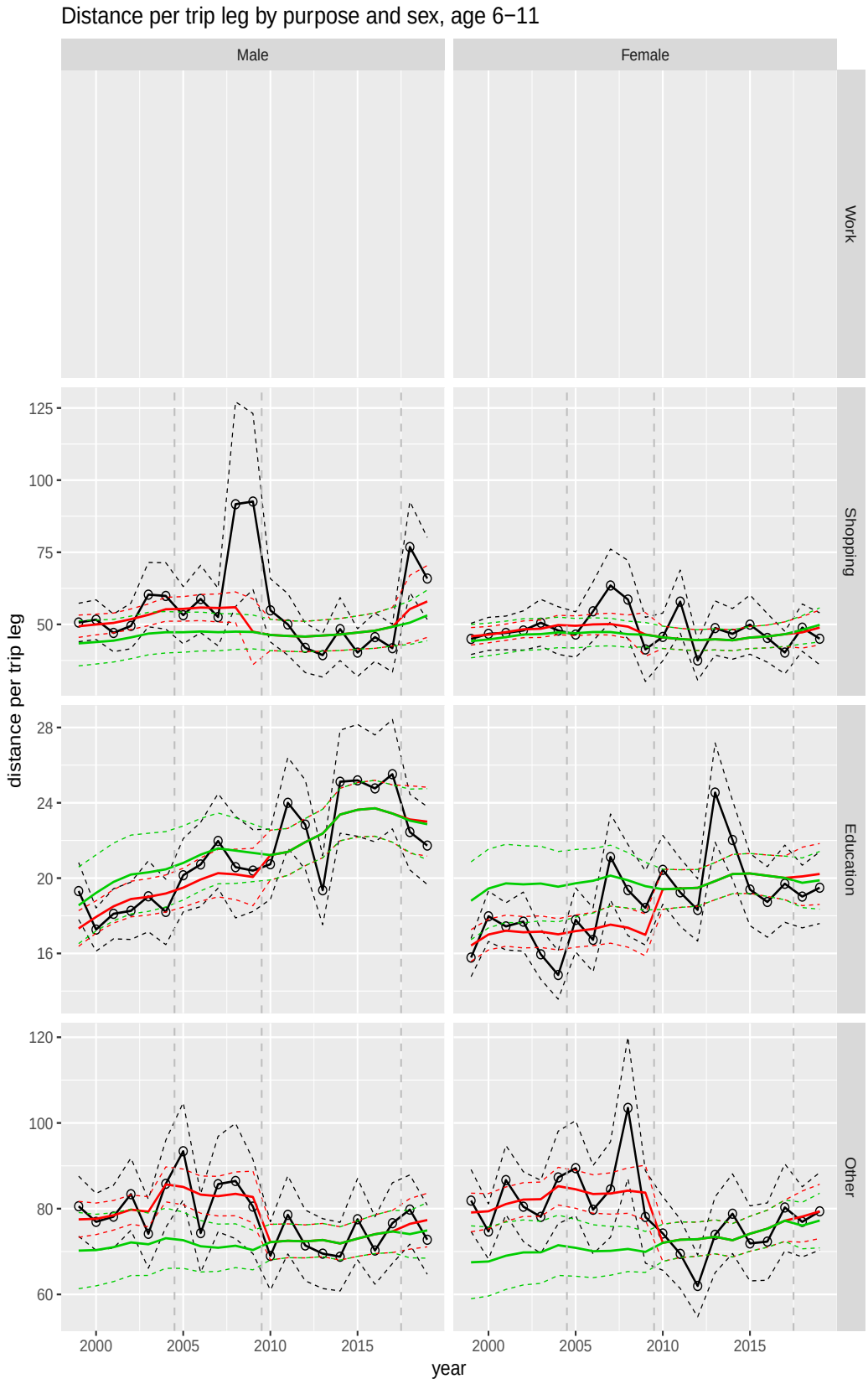

Figure A.76 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

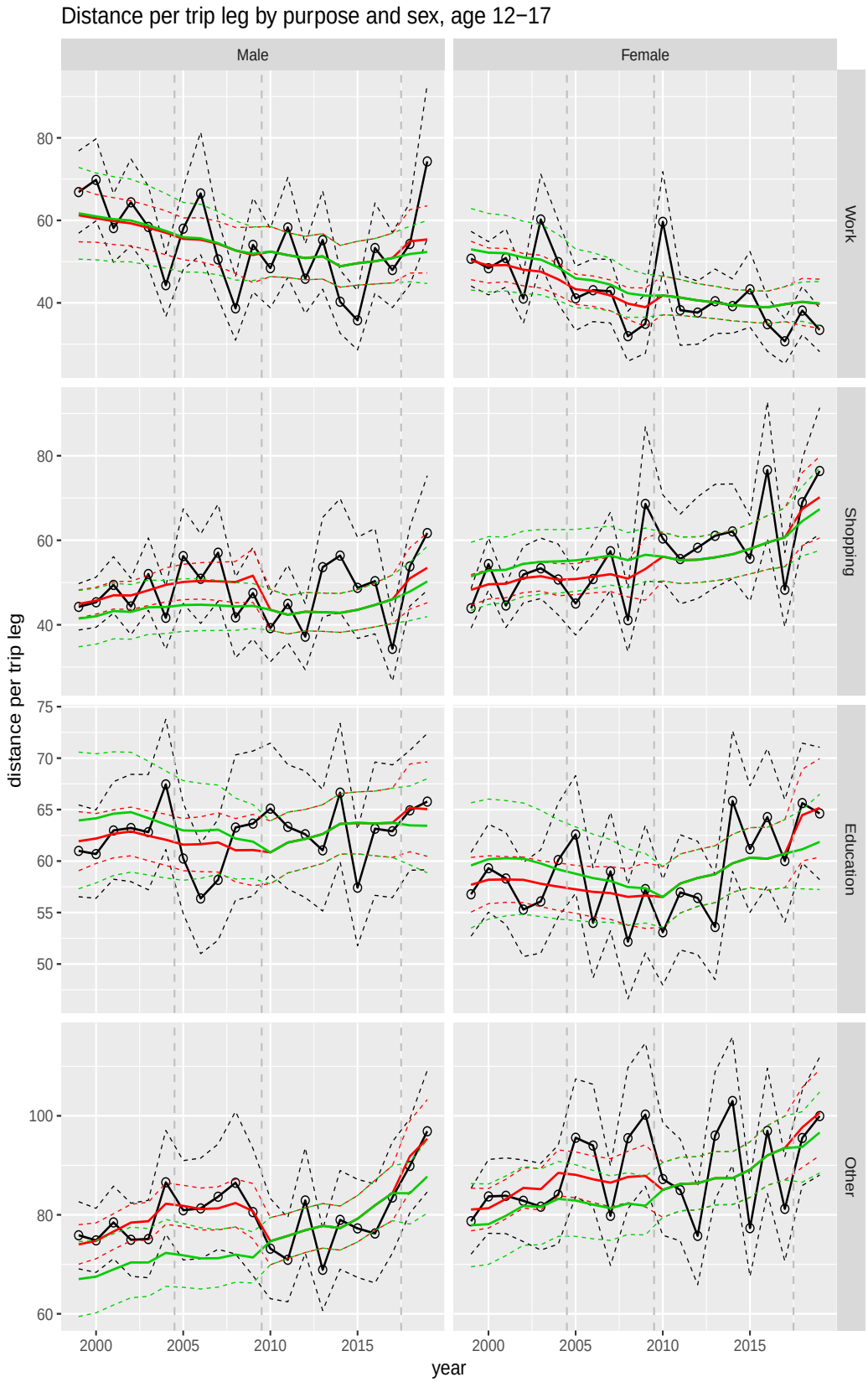

Figure A.77 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

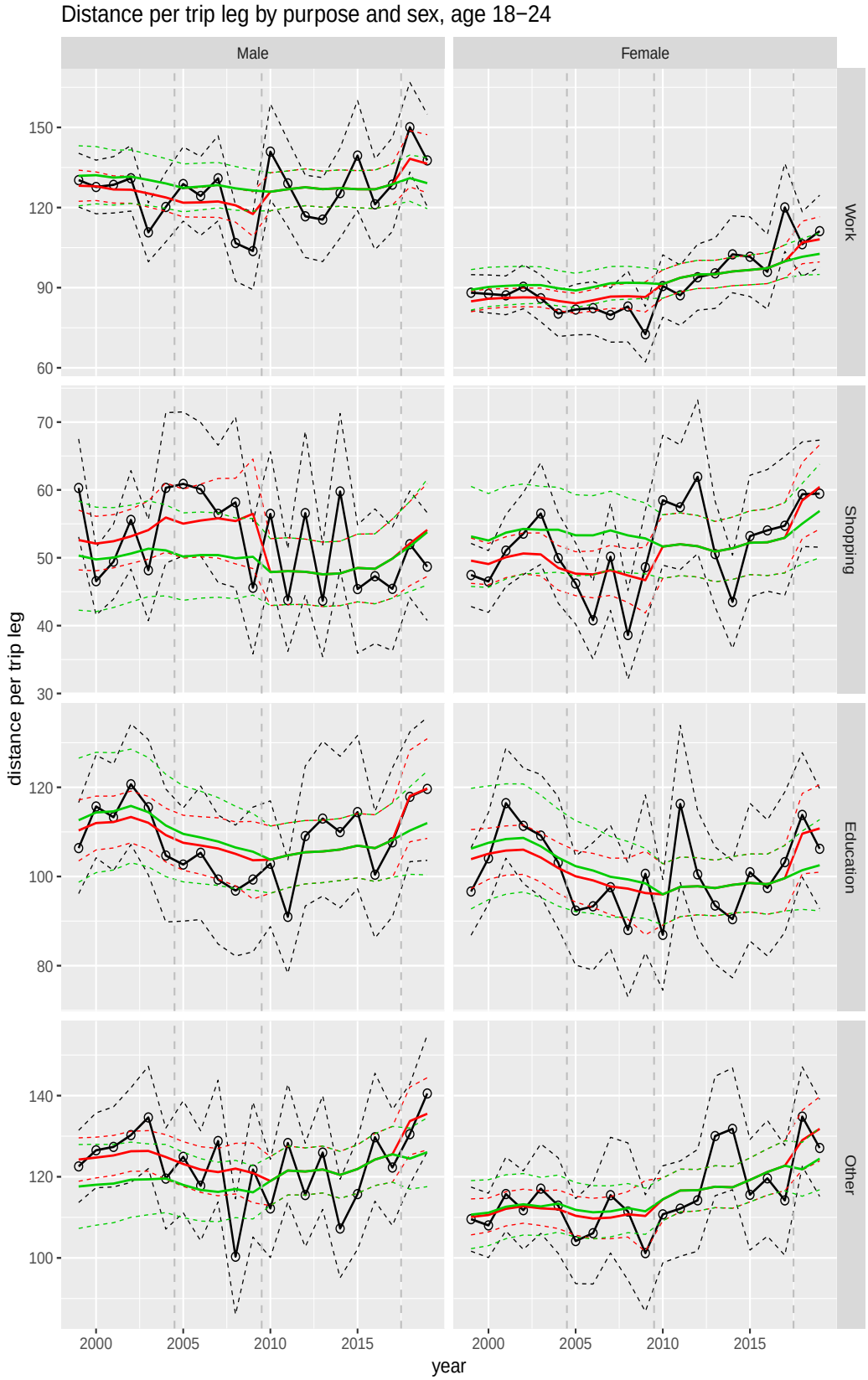

Figure A.78 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

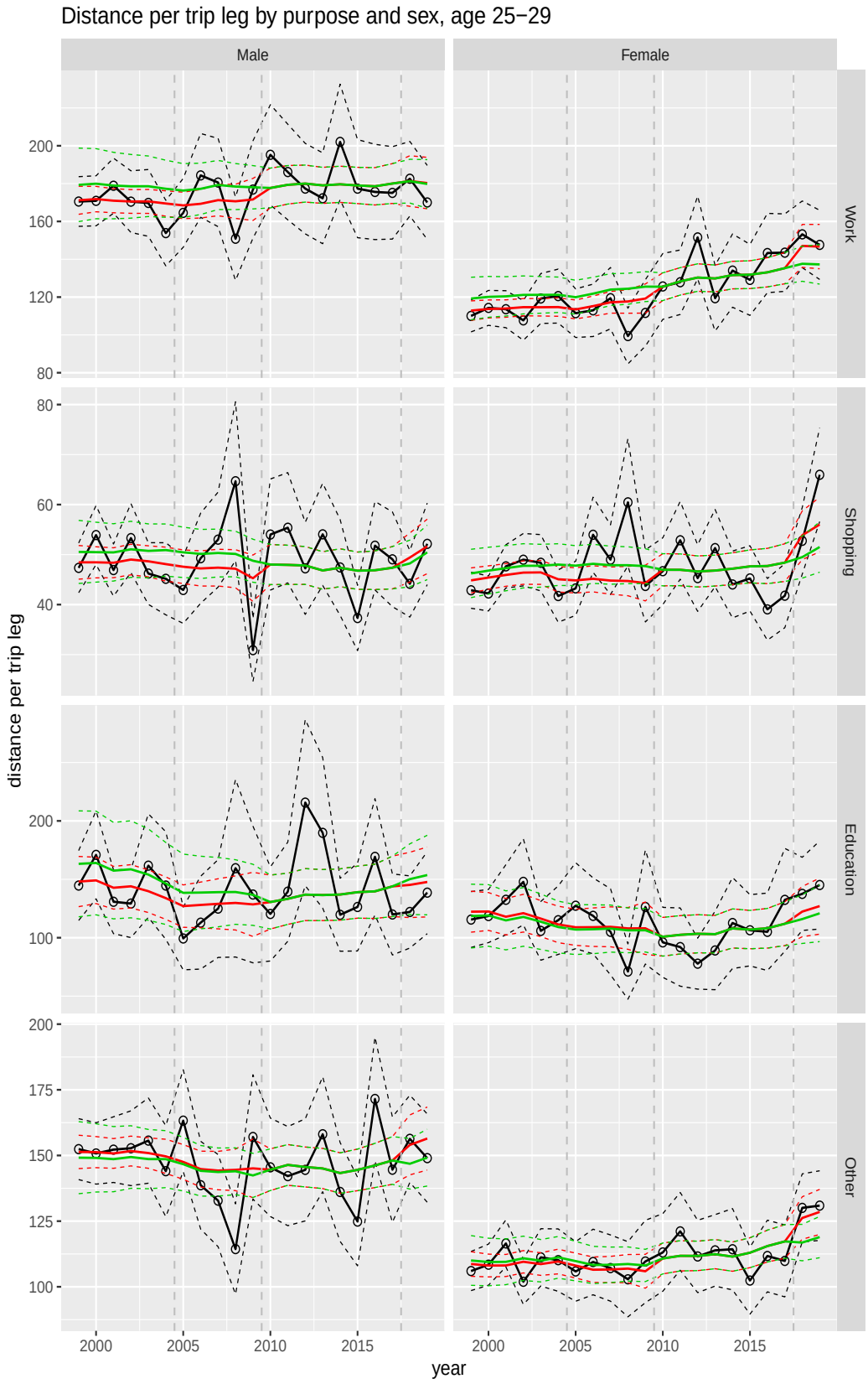

Figure A.79 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

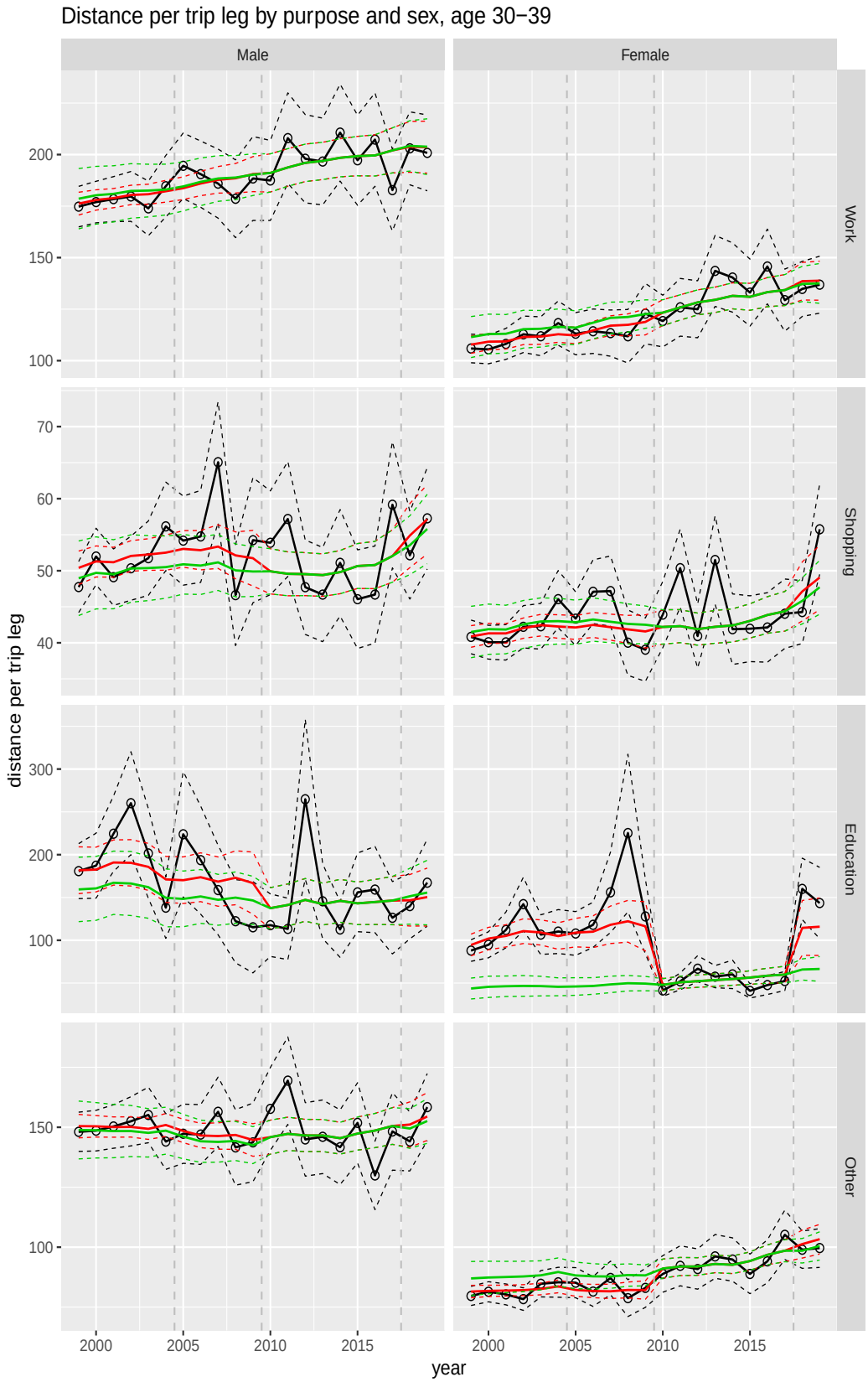

Figure A.80 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

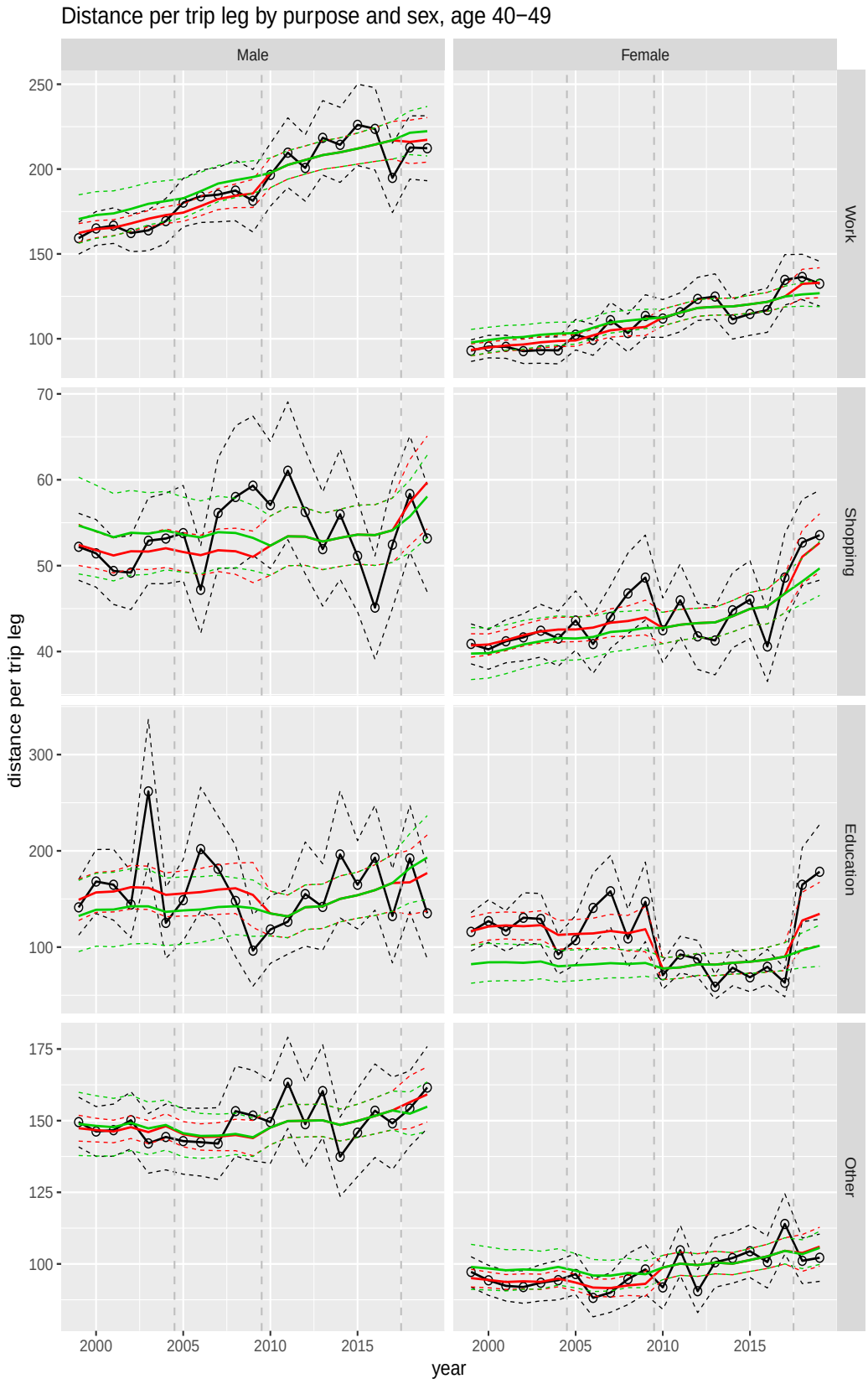

Figure A.81 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

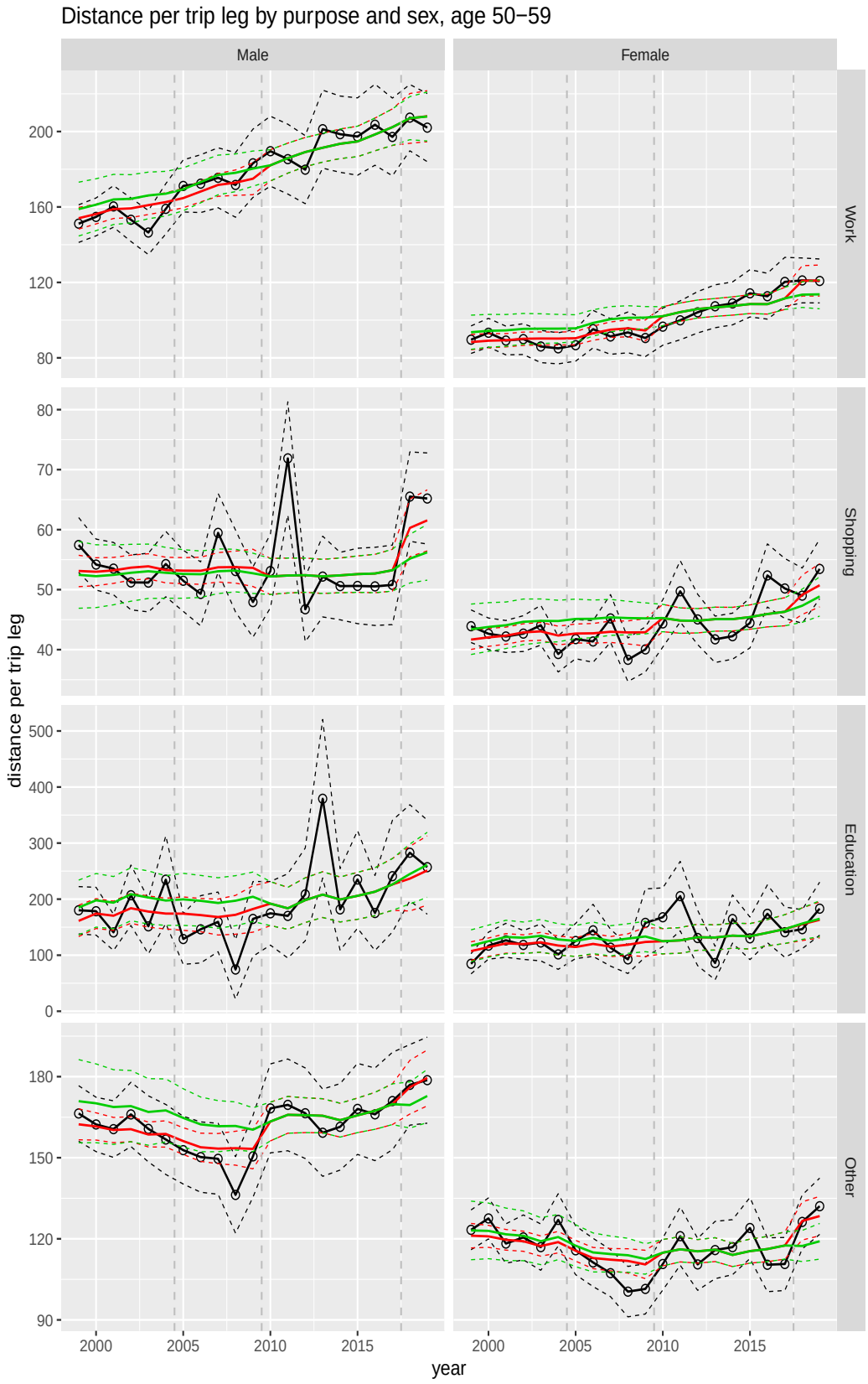

Figure A.82 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

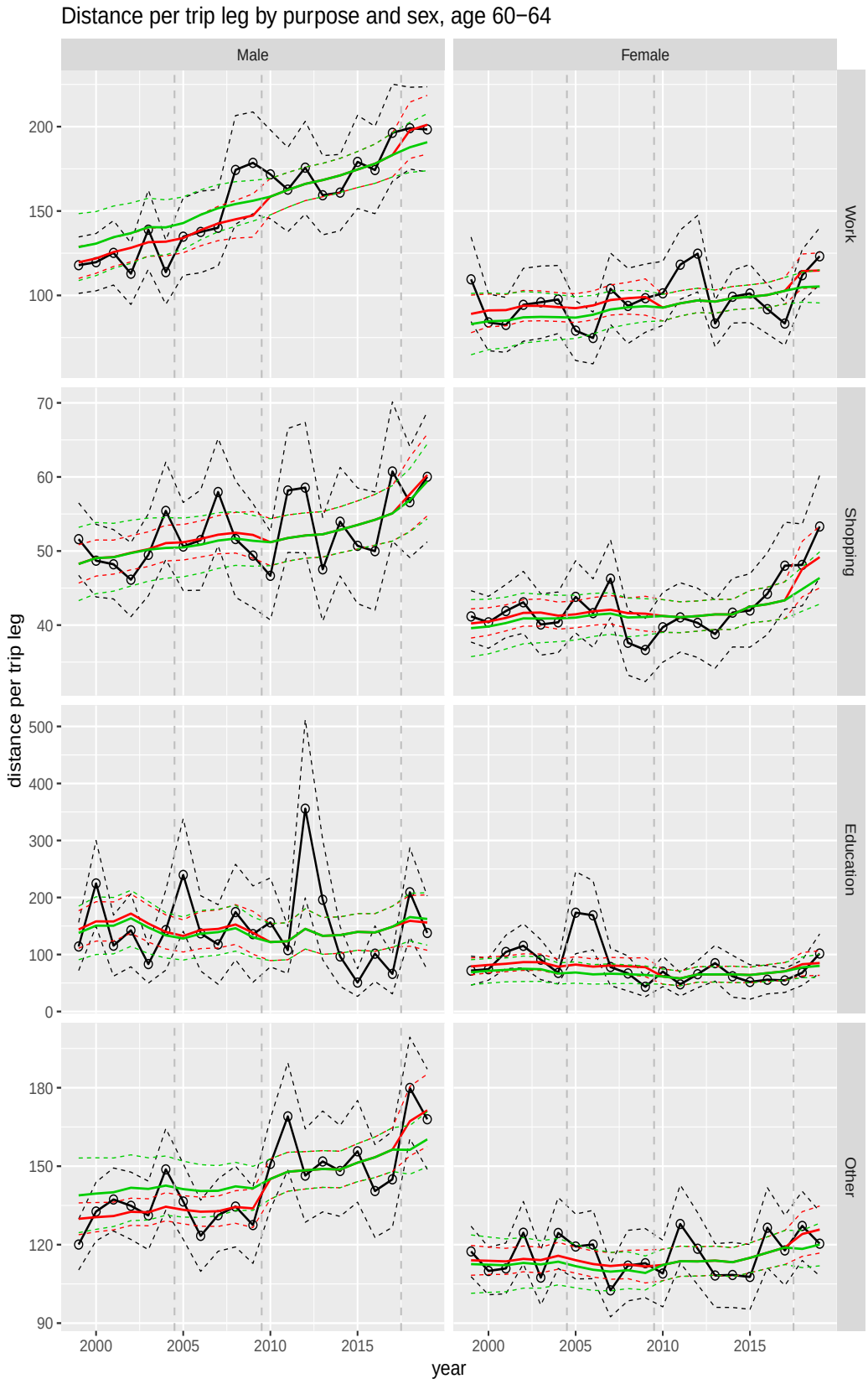

Figure A.83 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

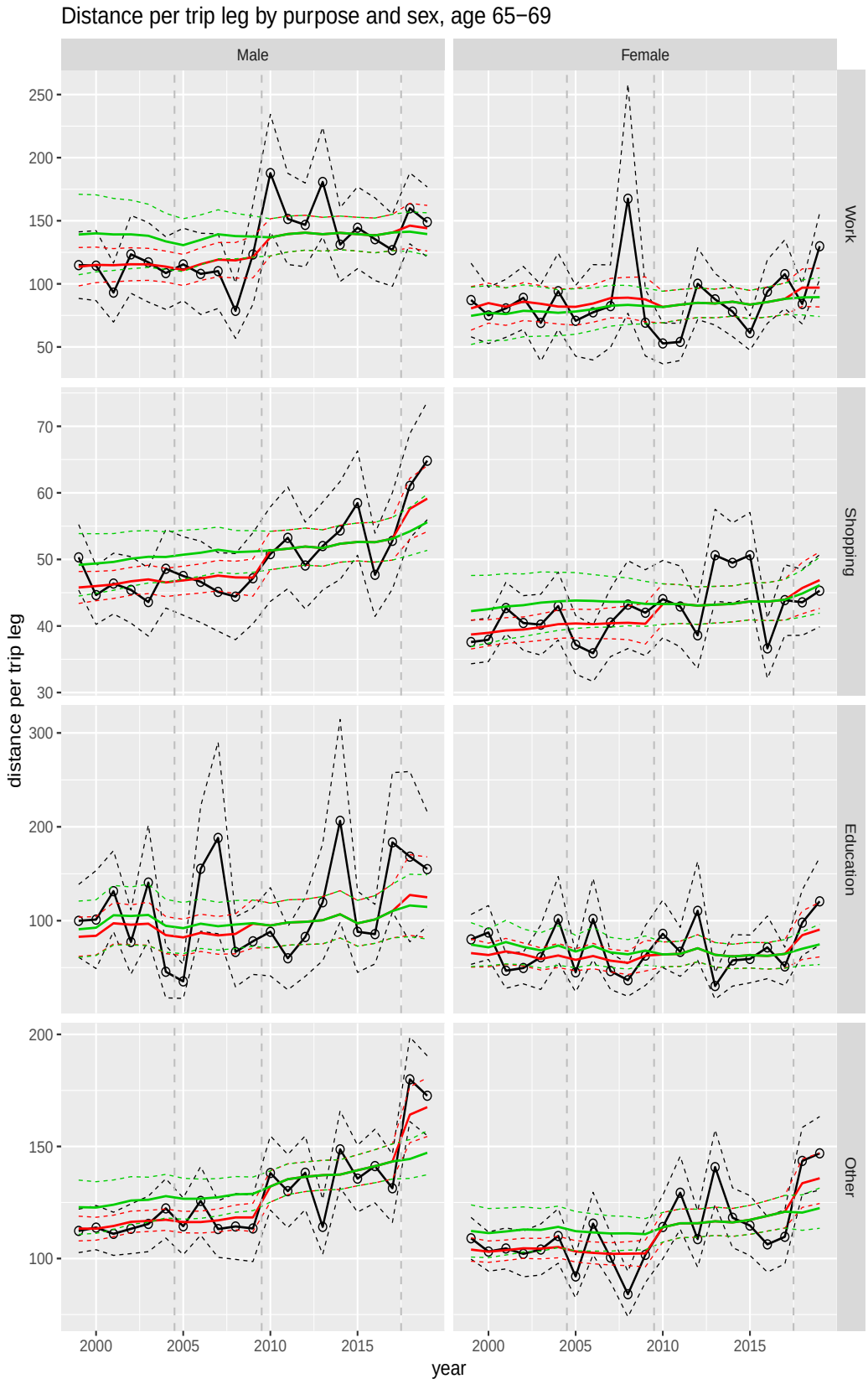

Figure A.84 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

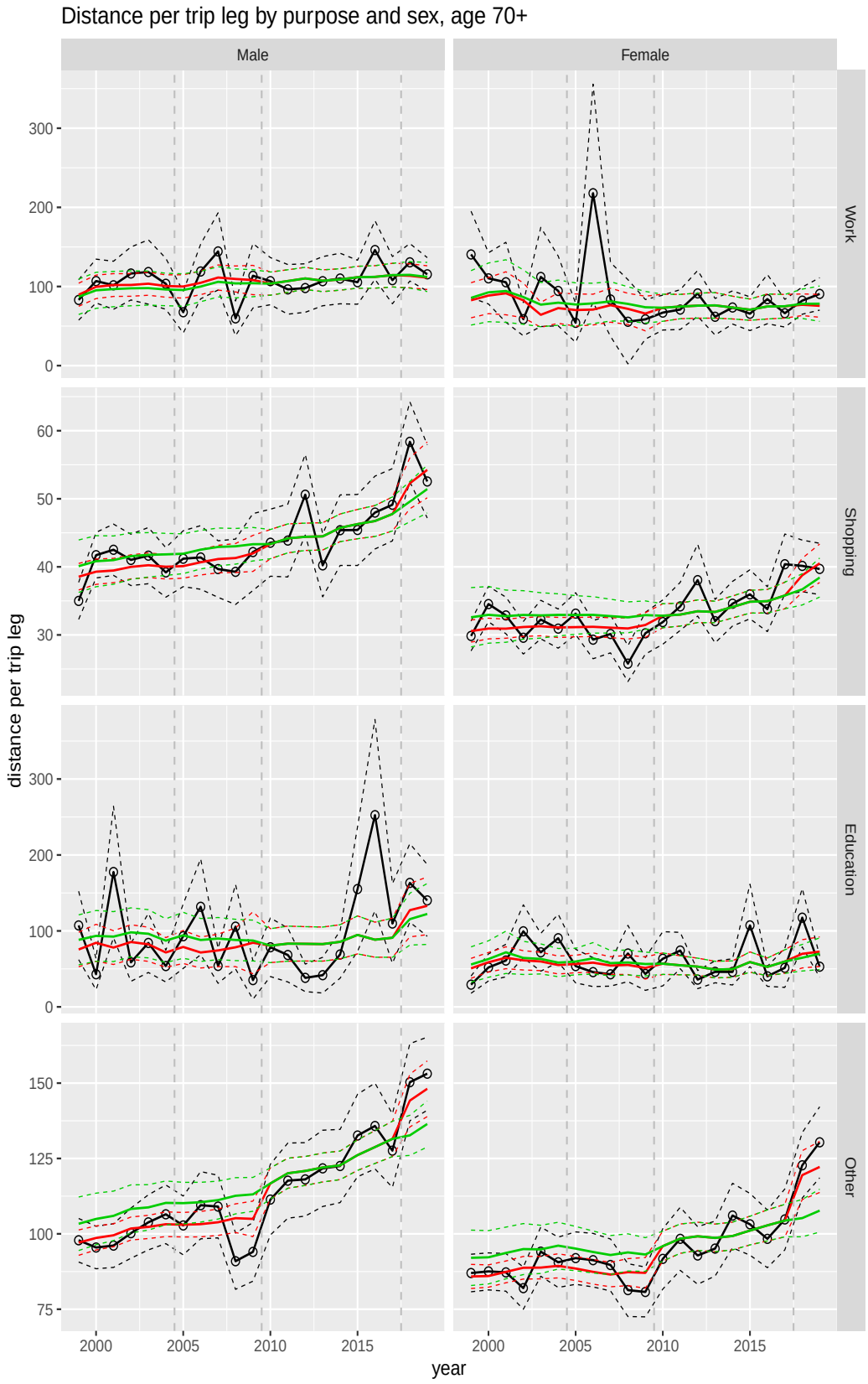

Figure A.85 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

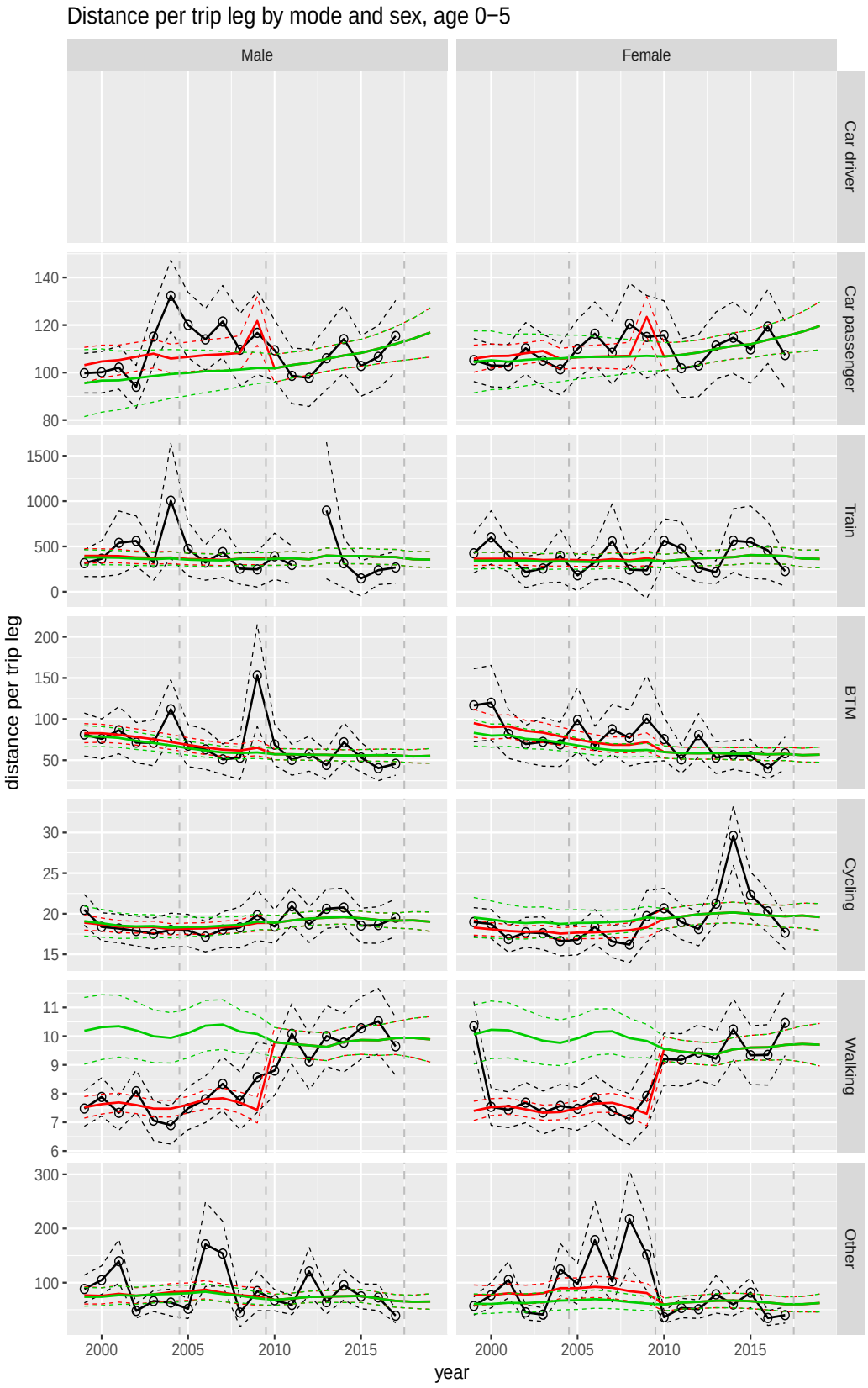

Figure A.86 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

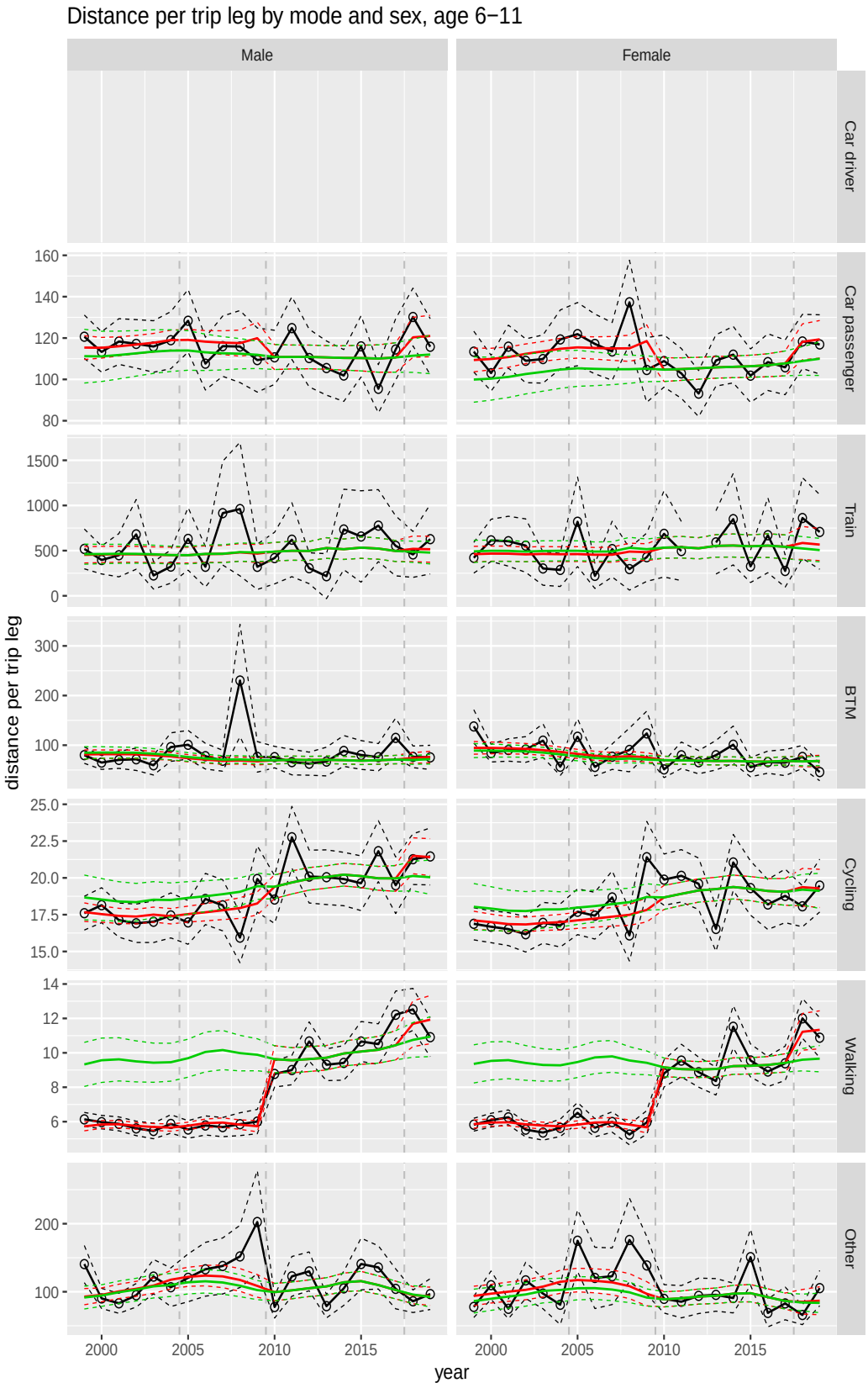

Figure A.87 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

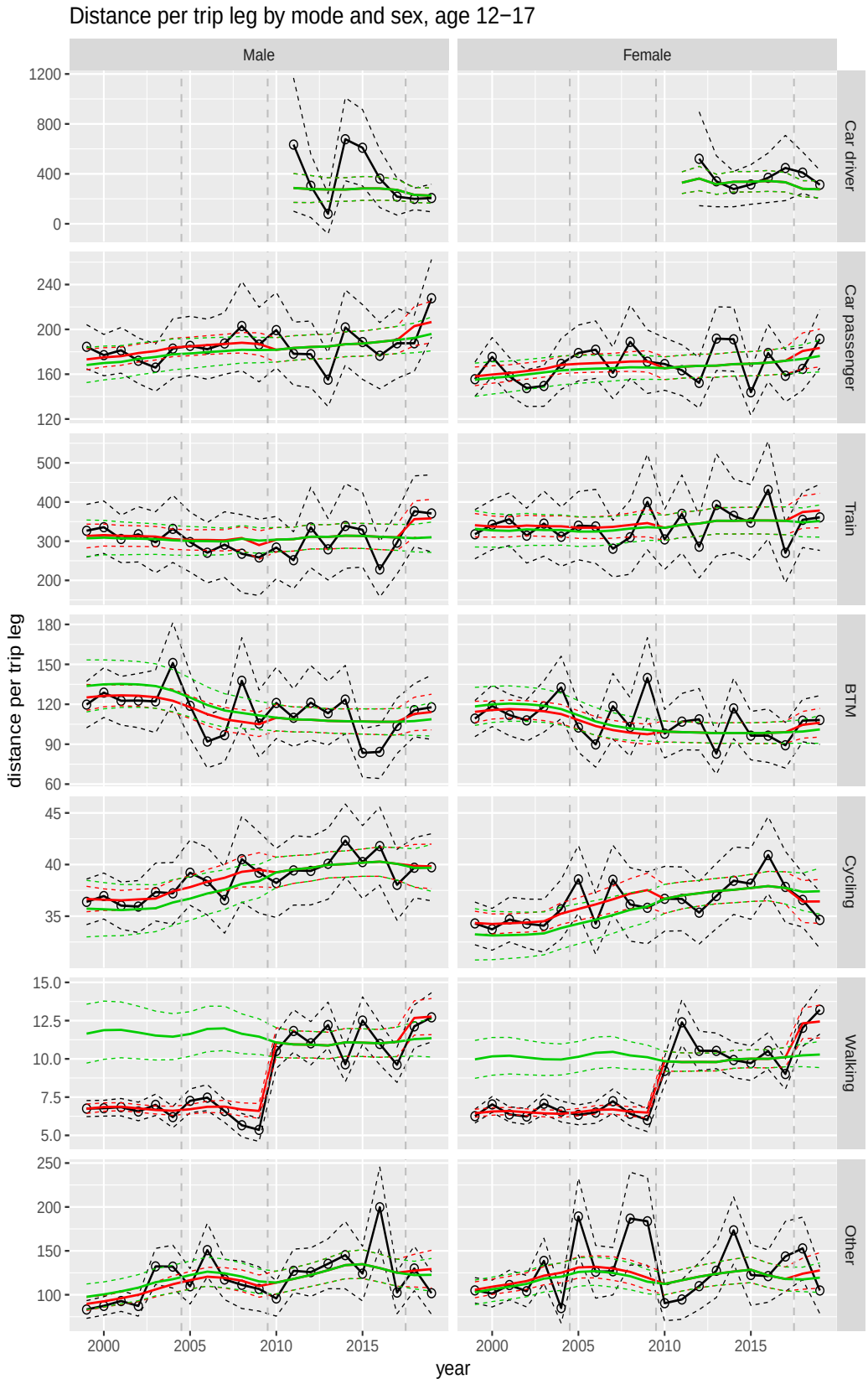

Figure A.88 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

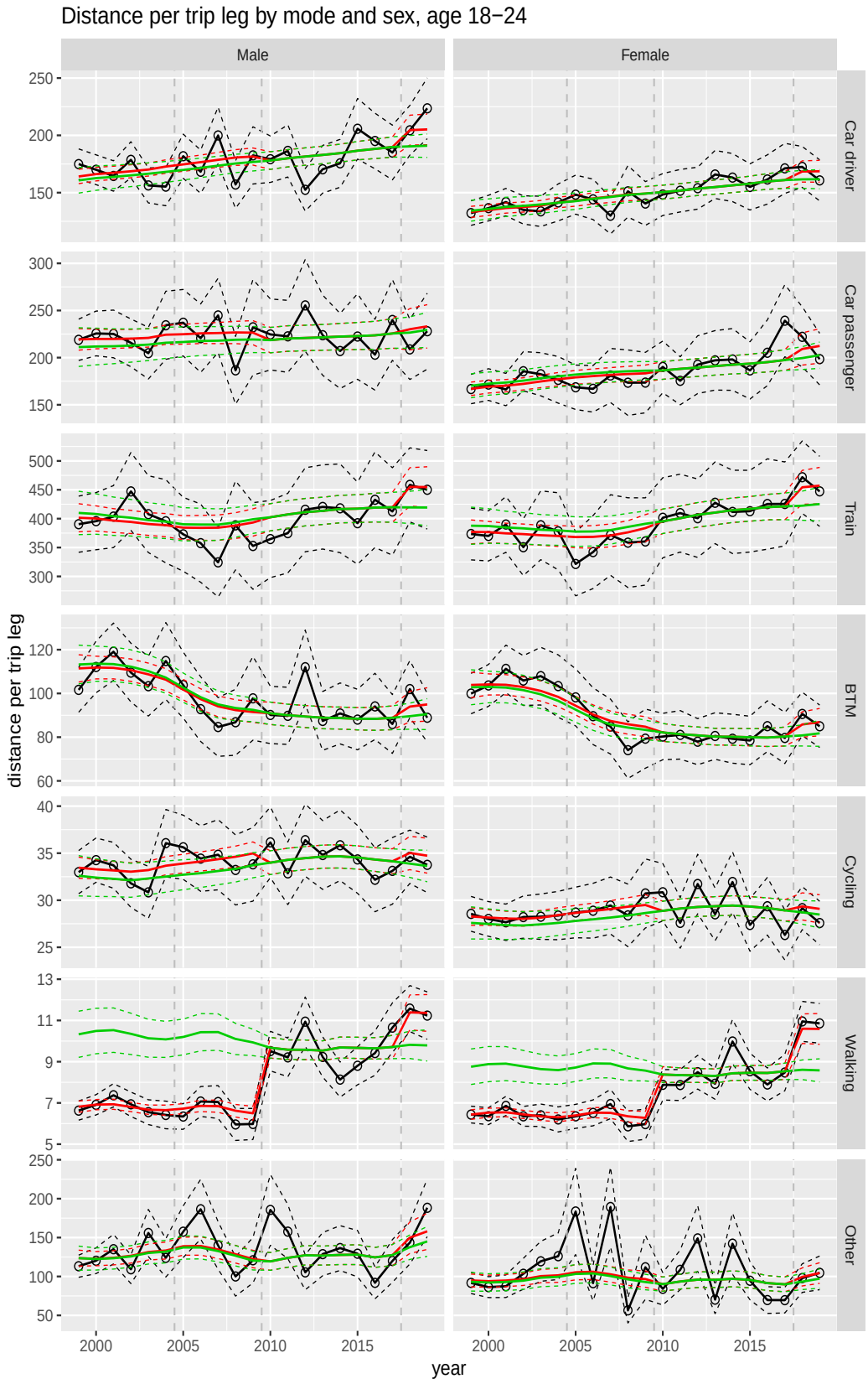

Figure A.89 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

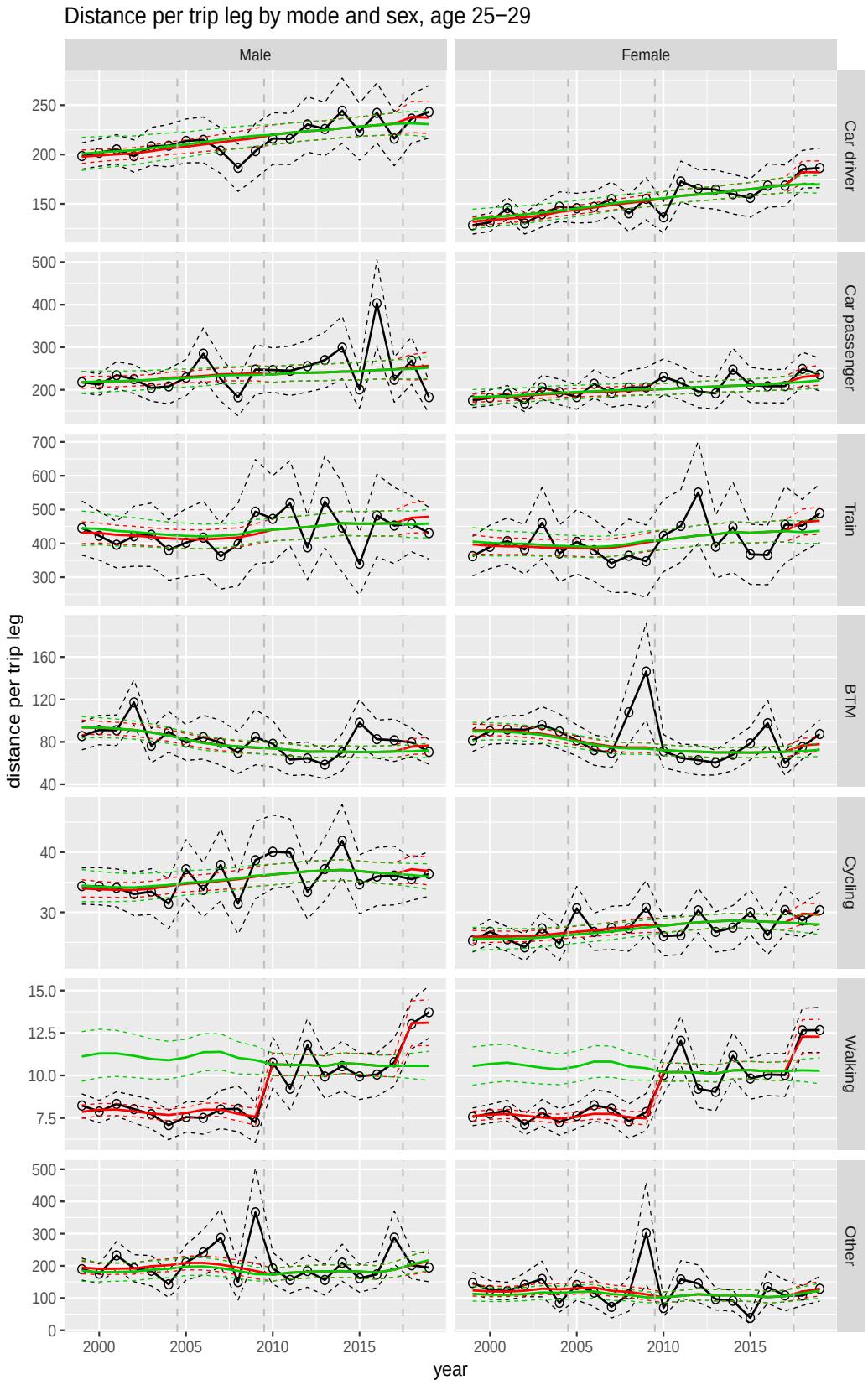

Figure A.90 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

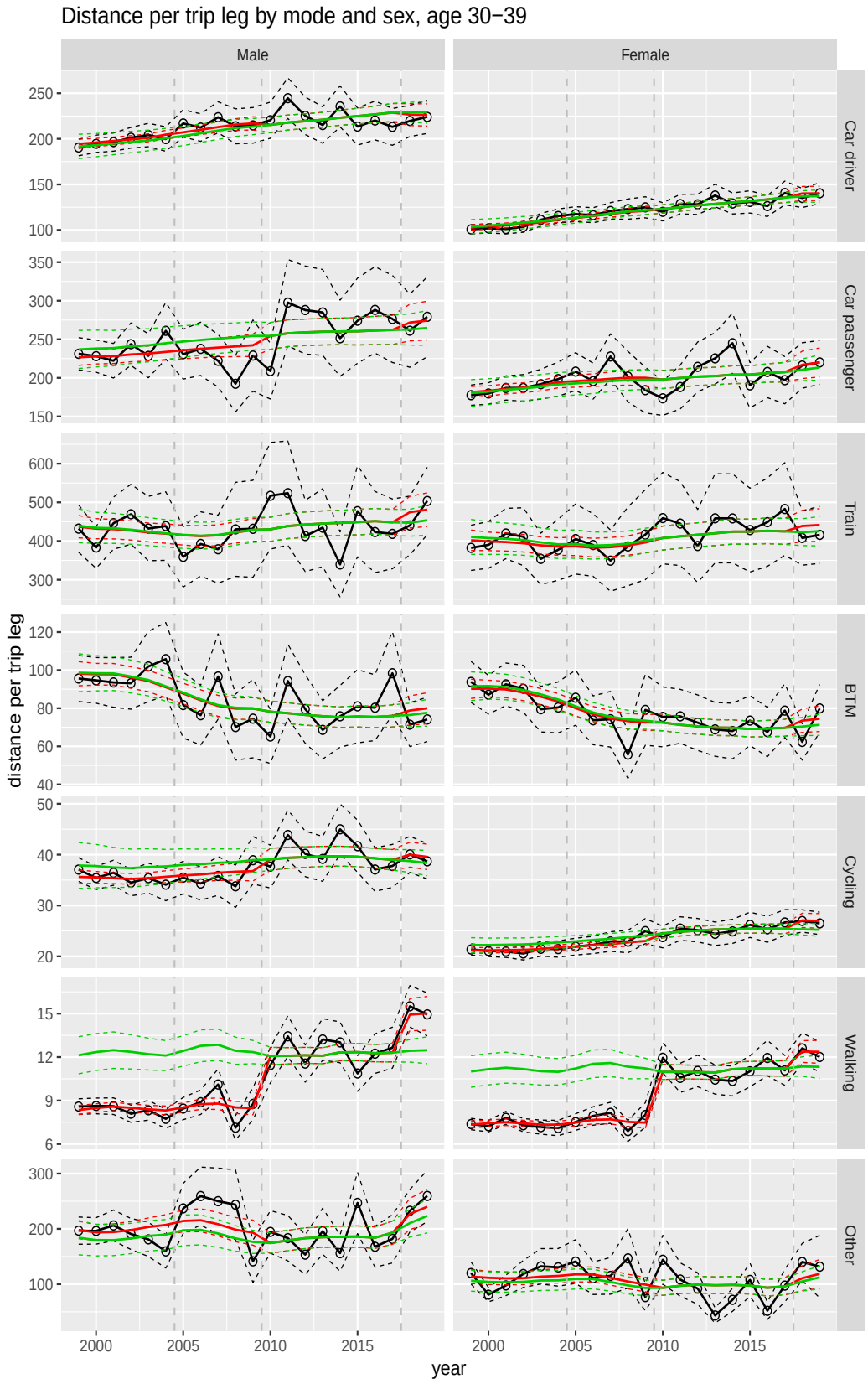

Figure A.91 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure A.92 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

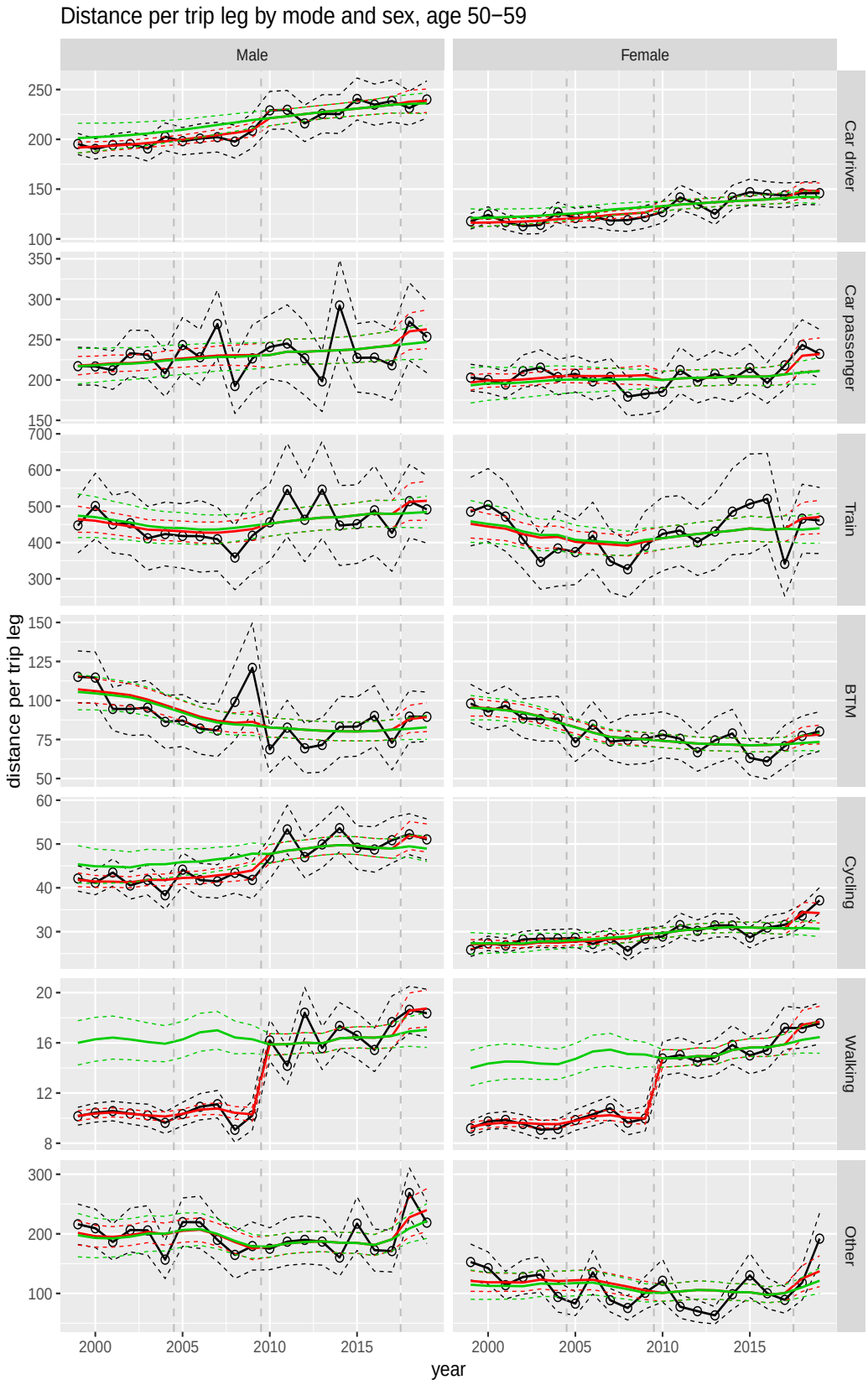

Figure A.93 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

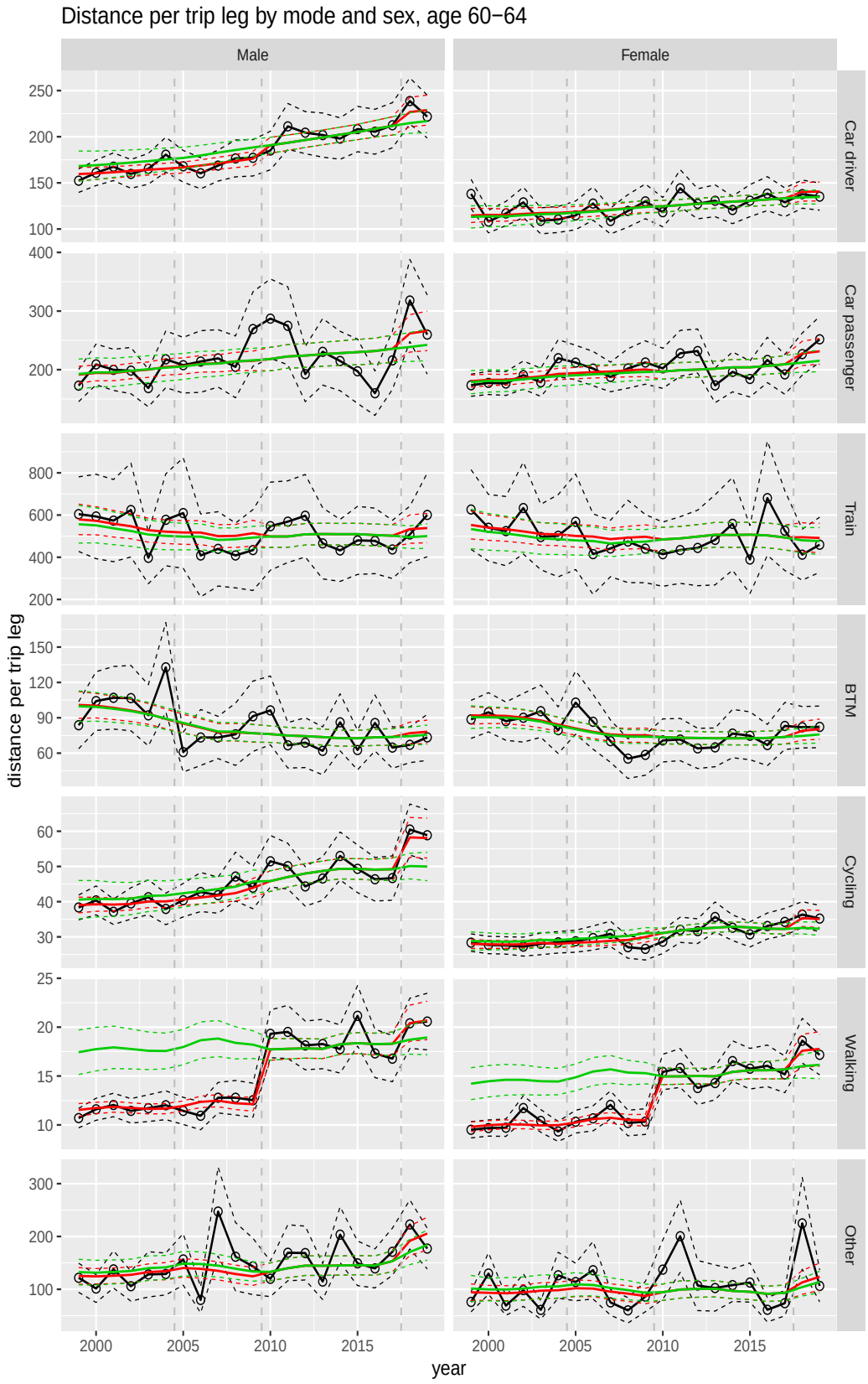

Figure A.94 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

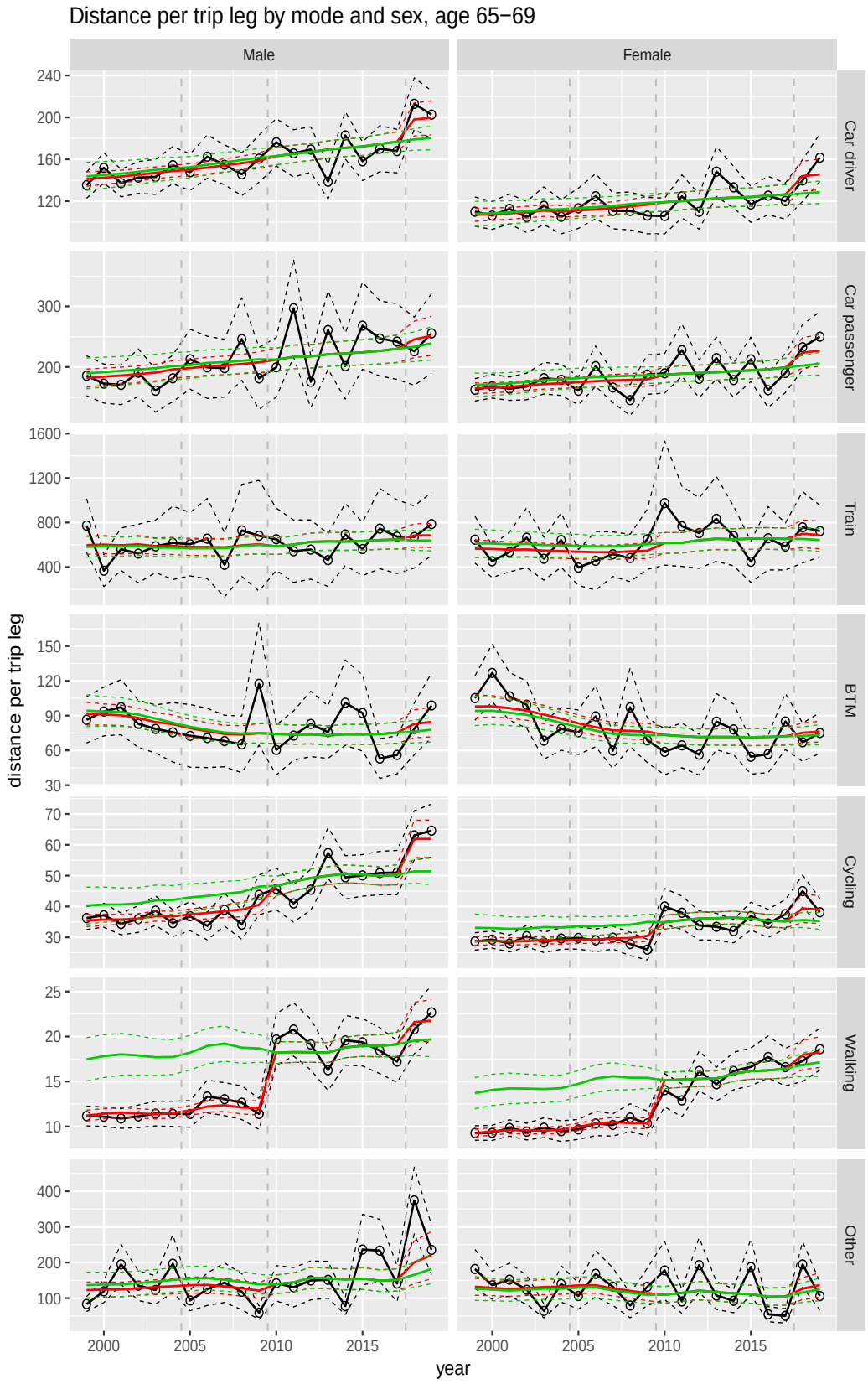

Figure A.95 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

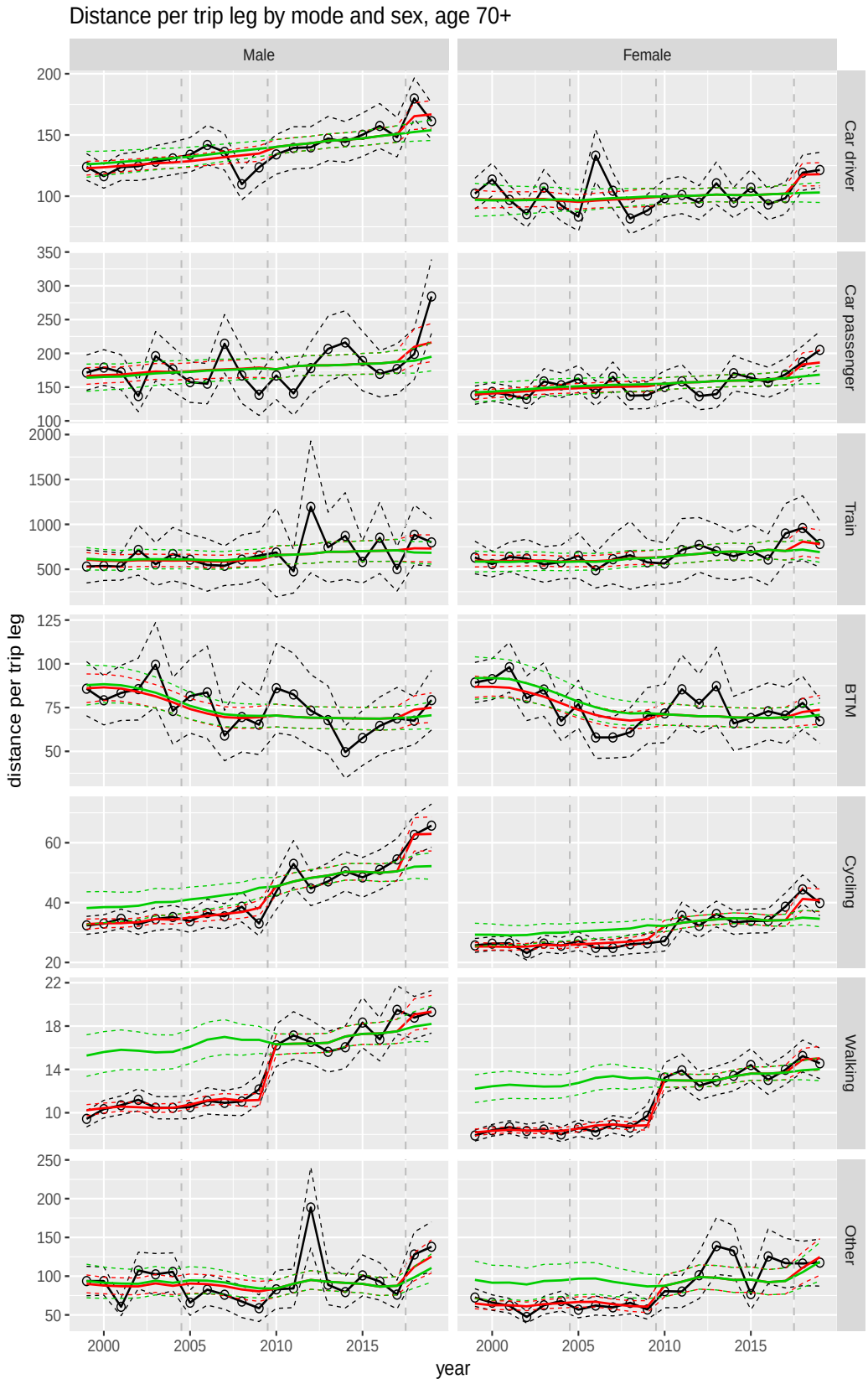

Figure A.96 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

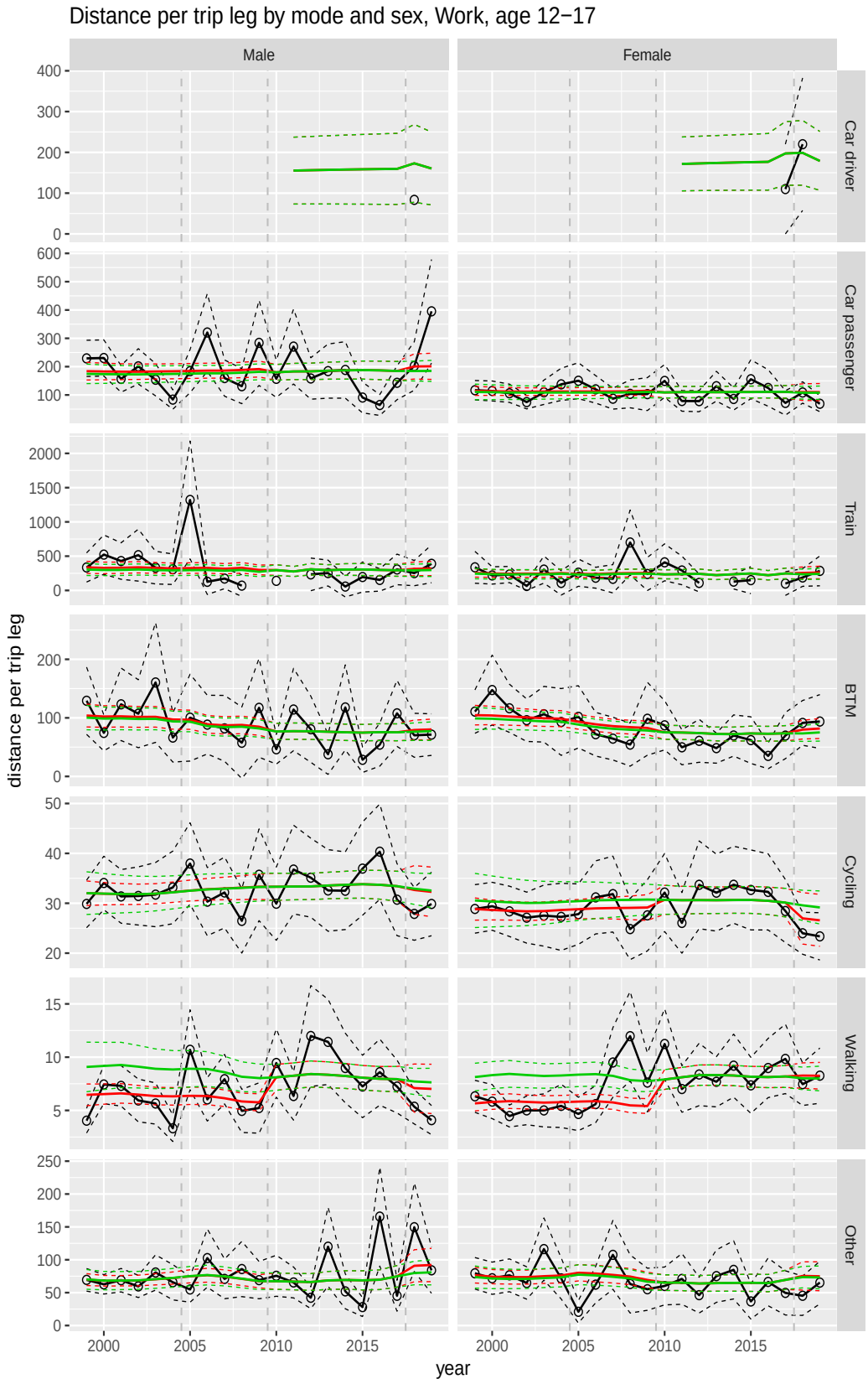

Figure A.97 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

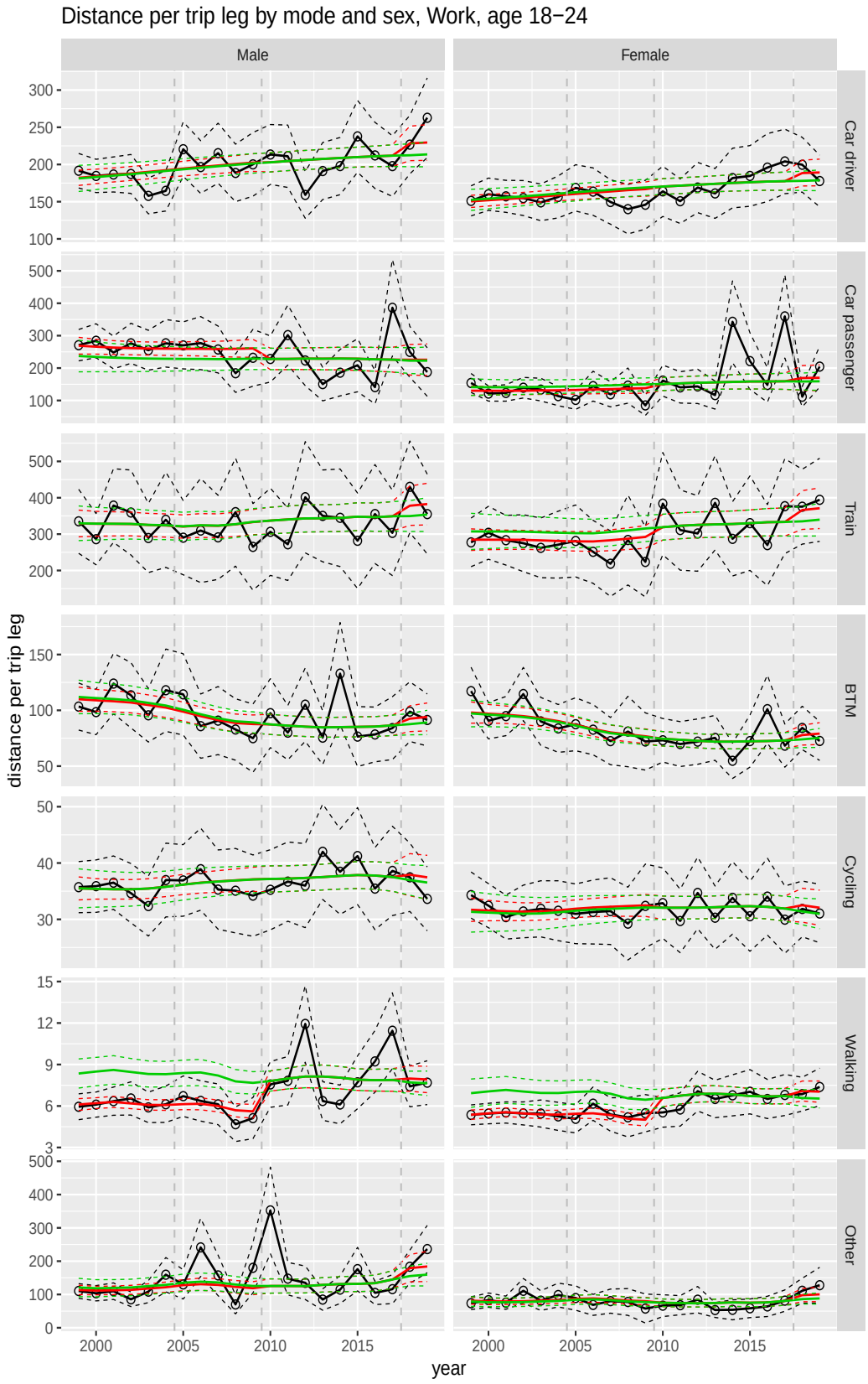

Figure A.98 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

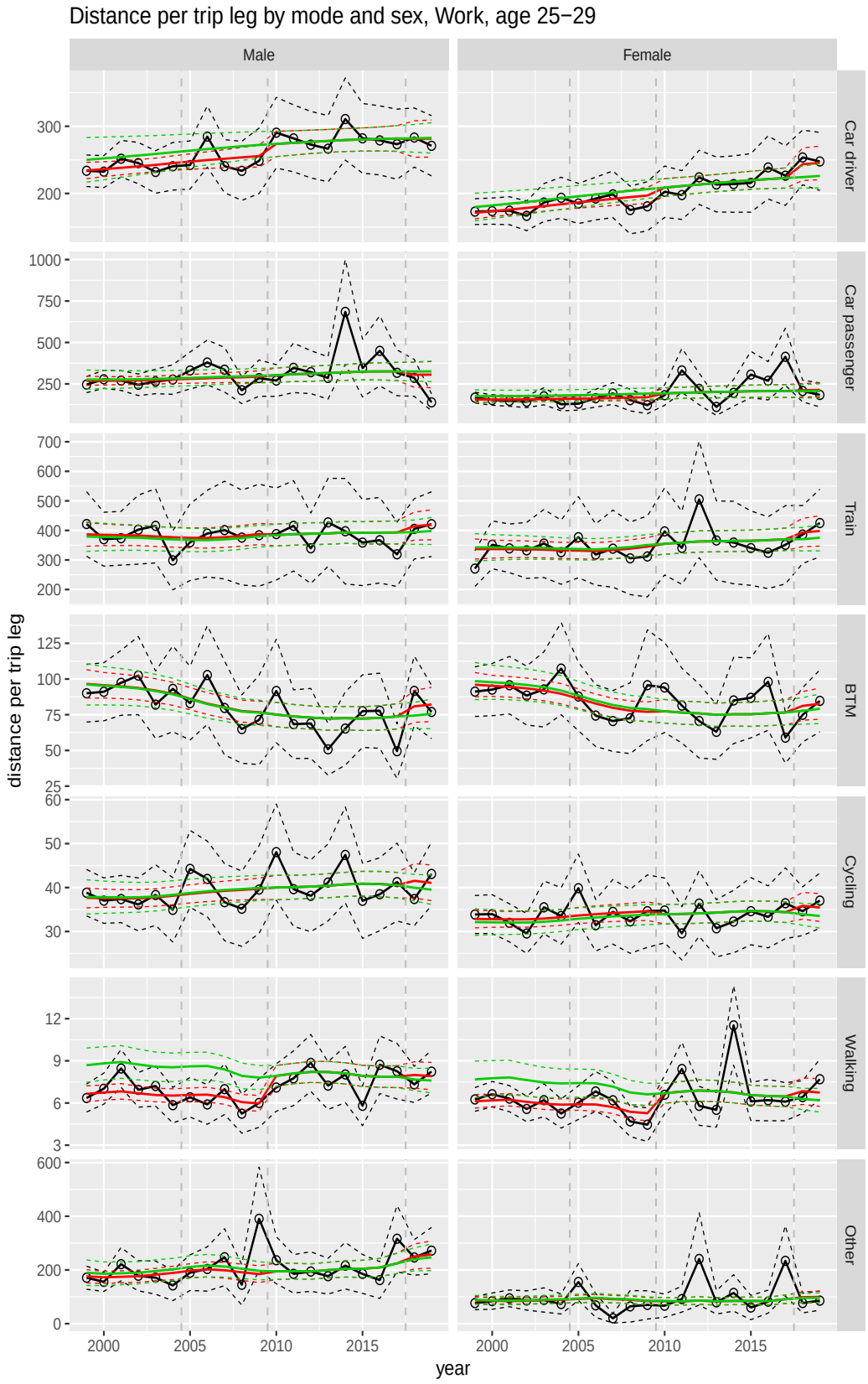

Figure A.99 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

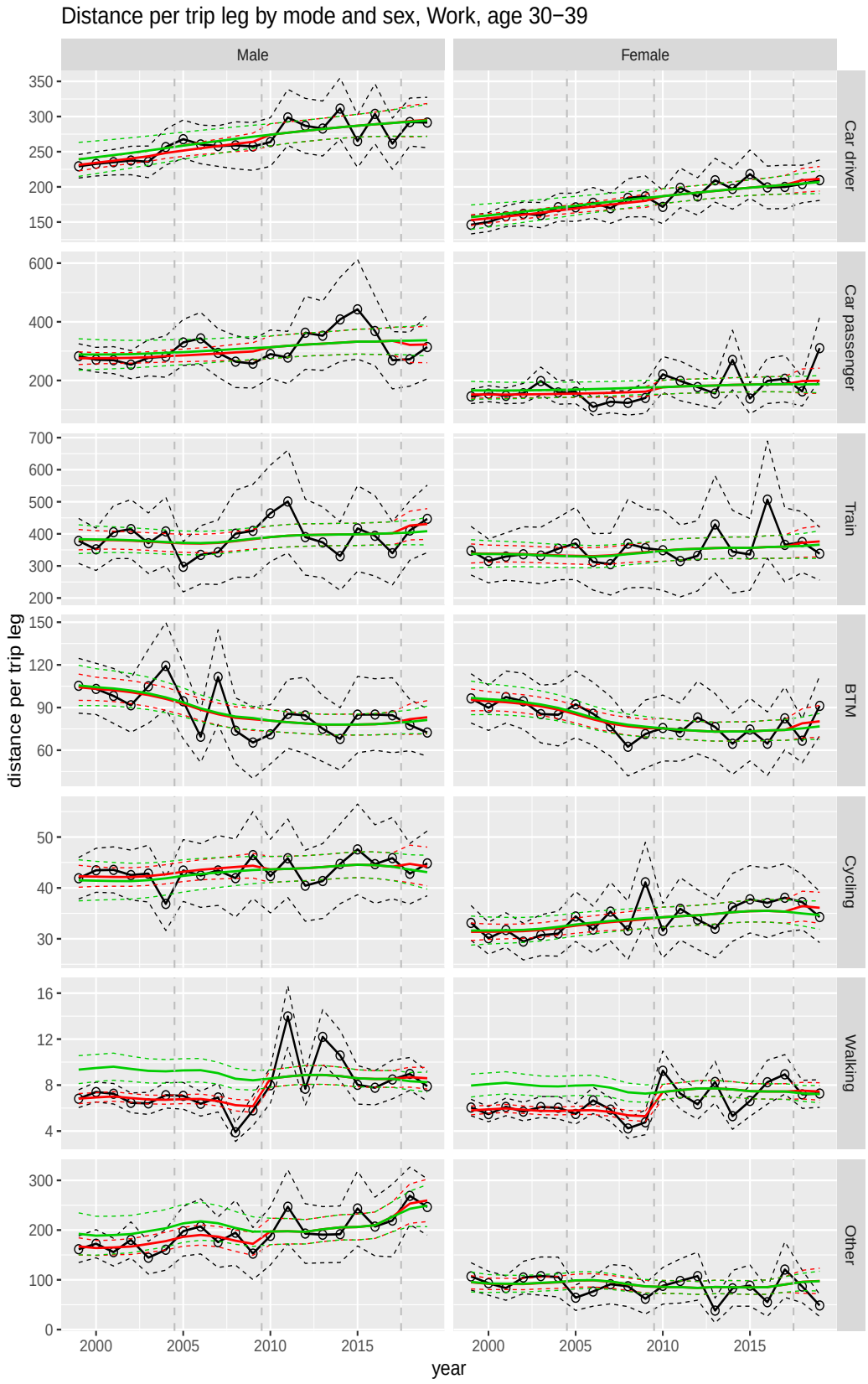

Figure A.100 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

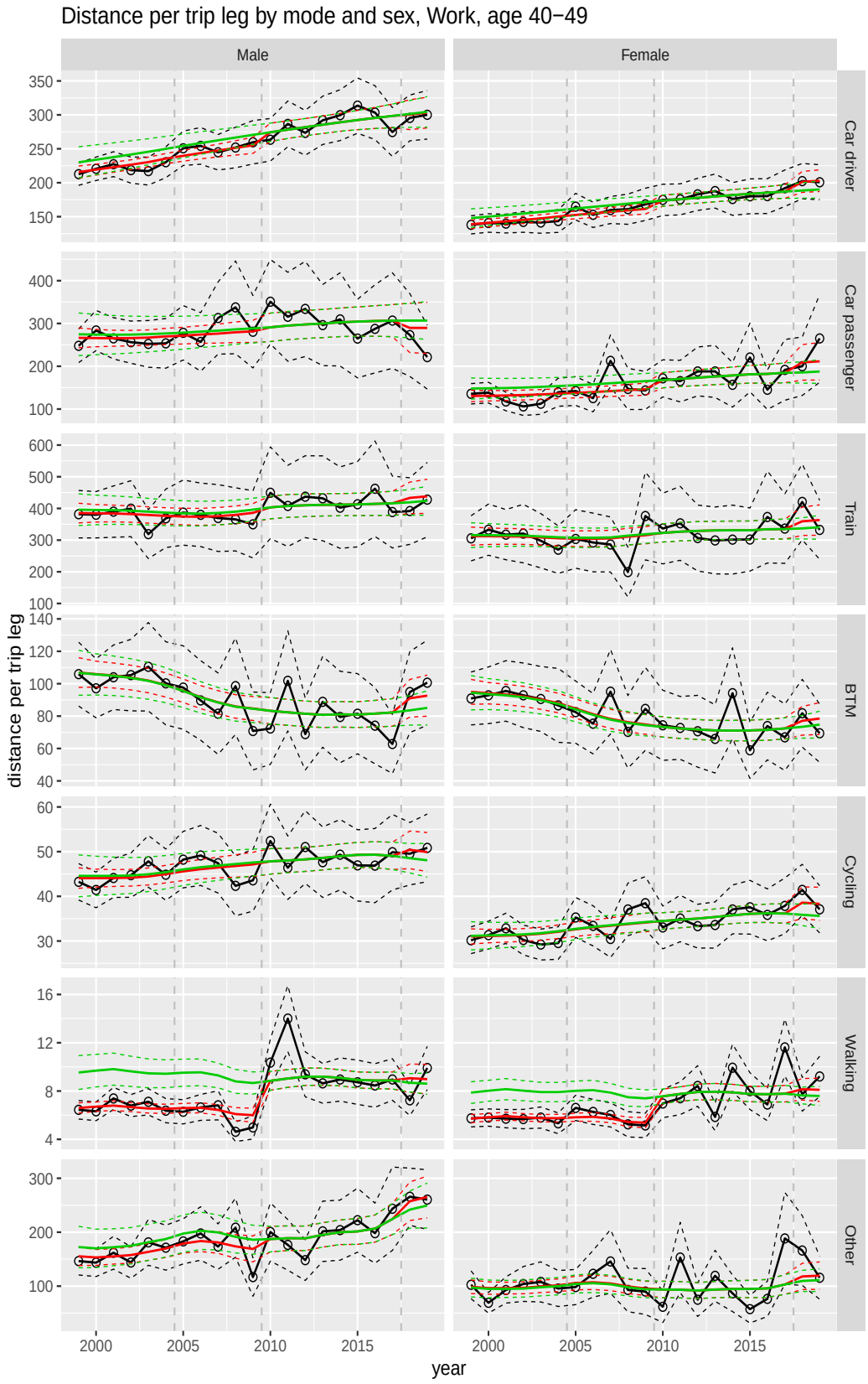

Figure A.101 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

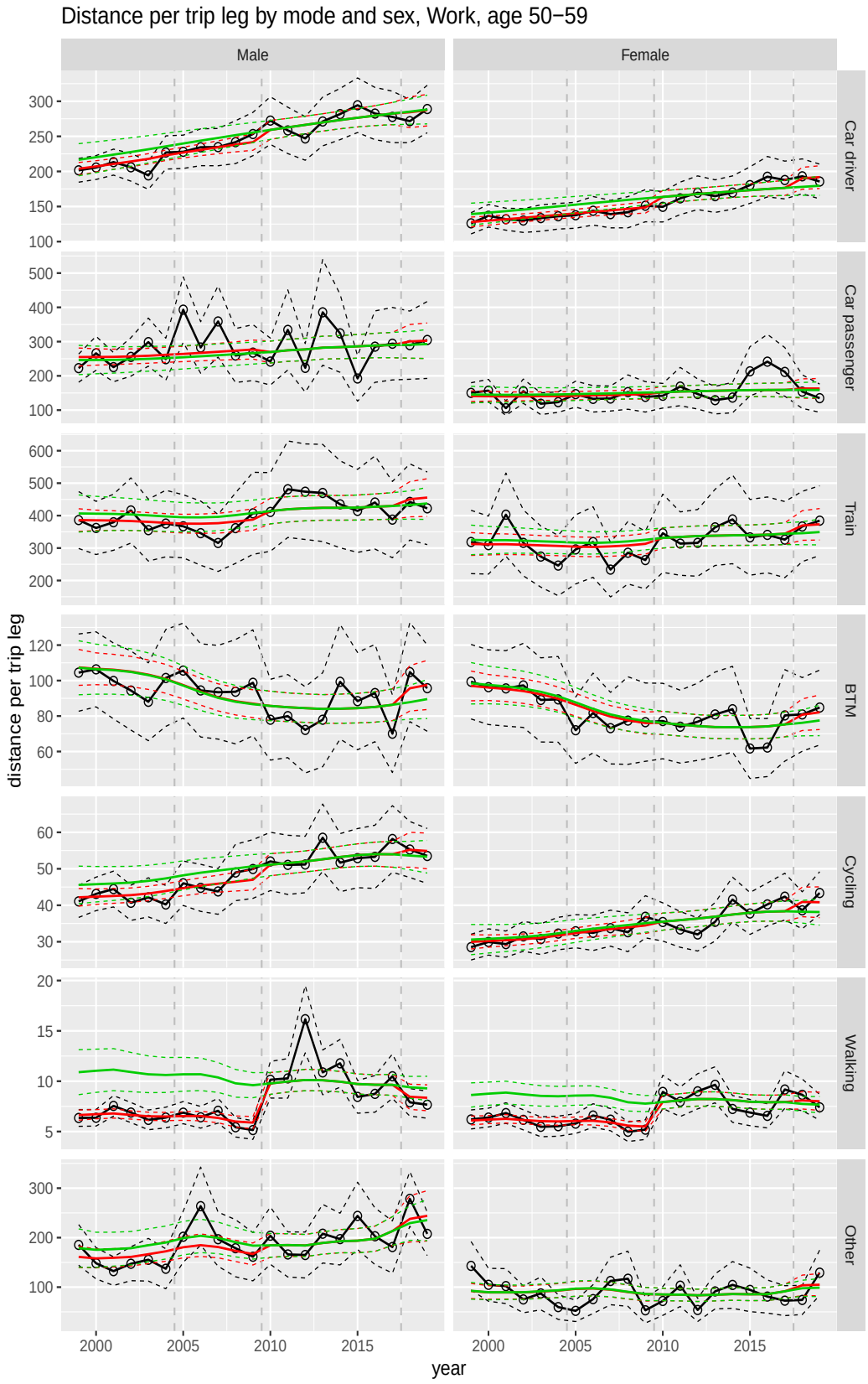

Figure A.102 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

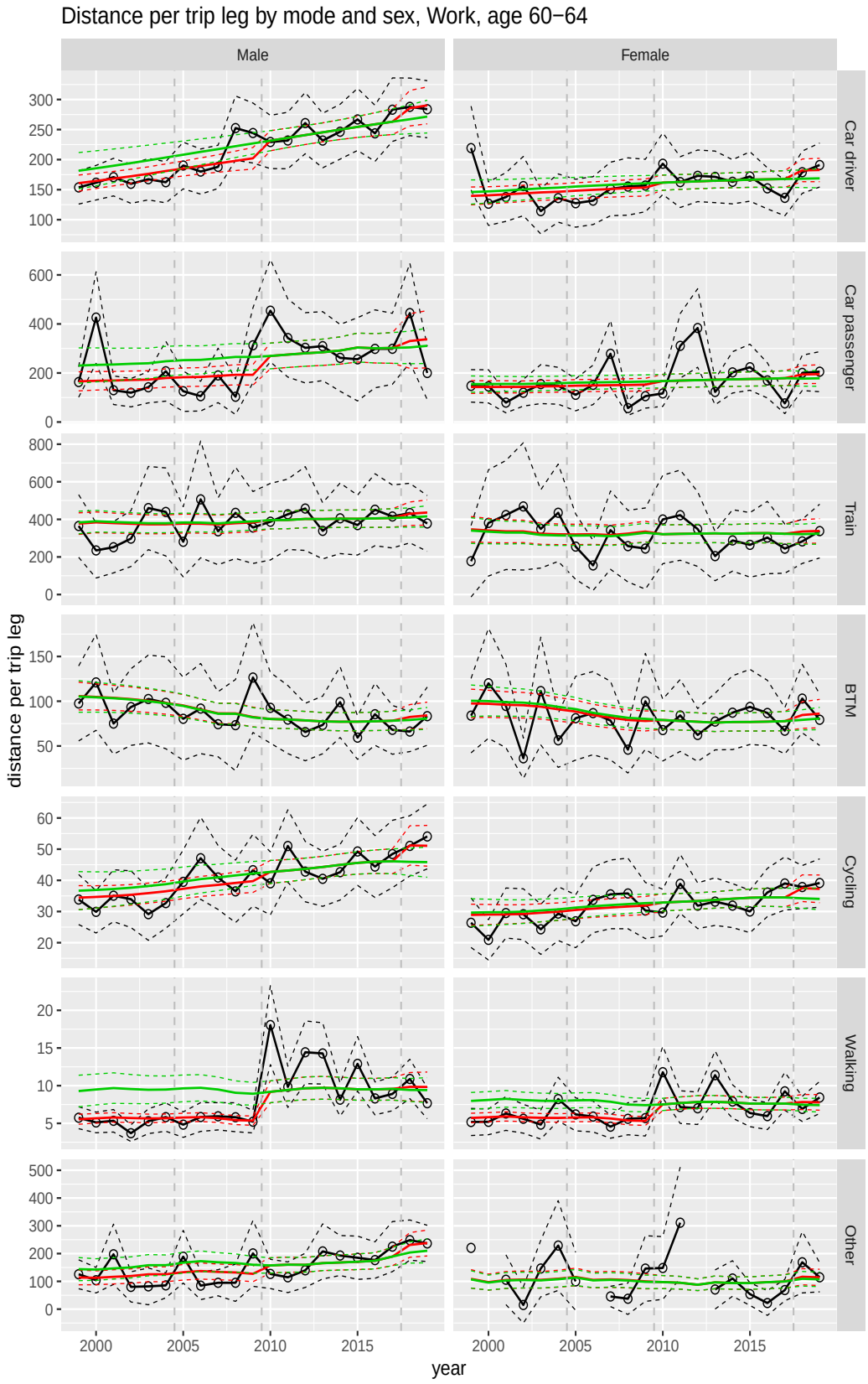

Figure A.103 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

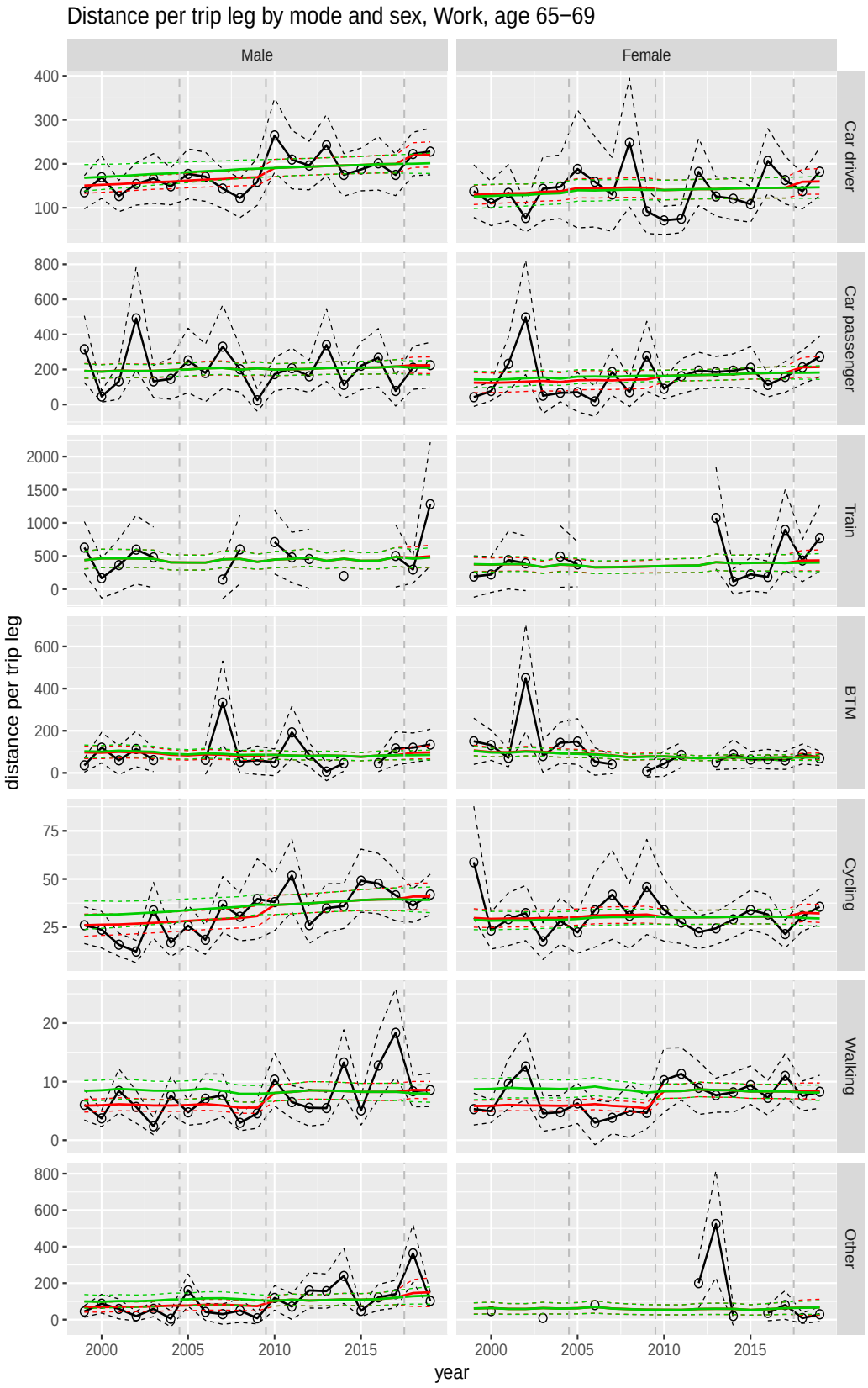

Figure A.104 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

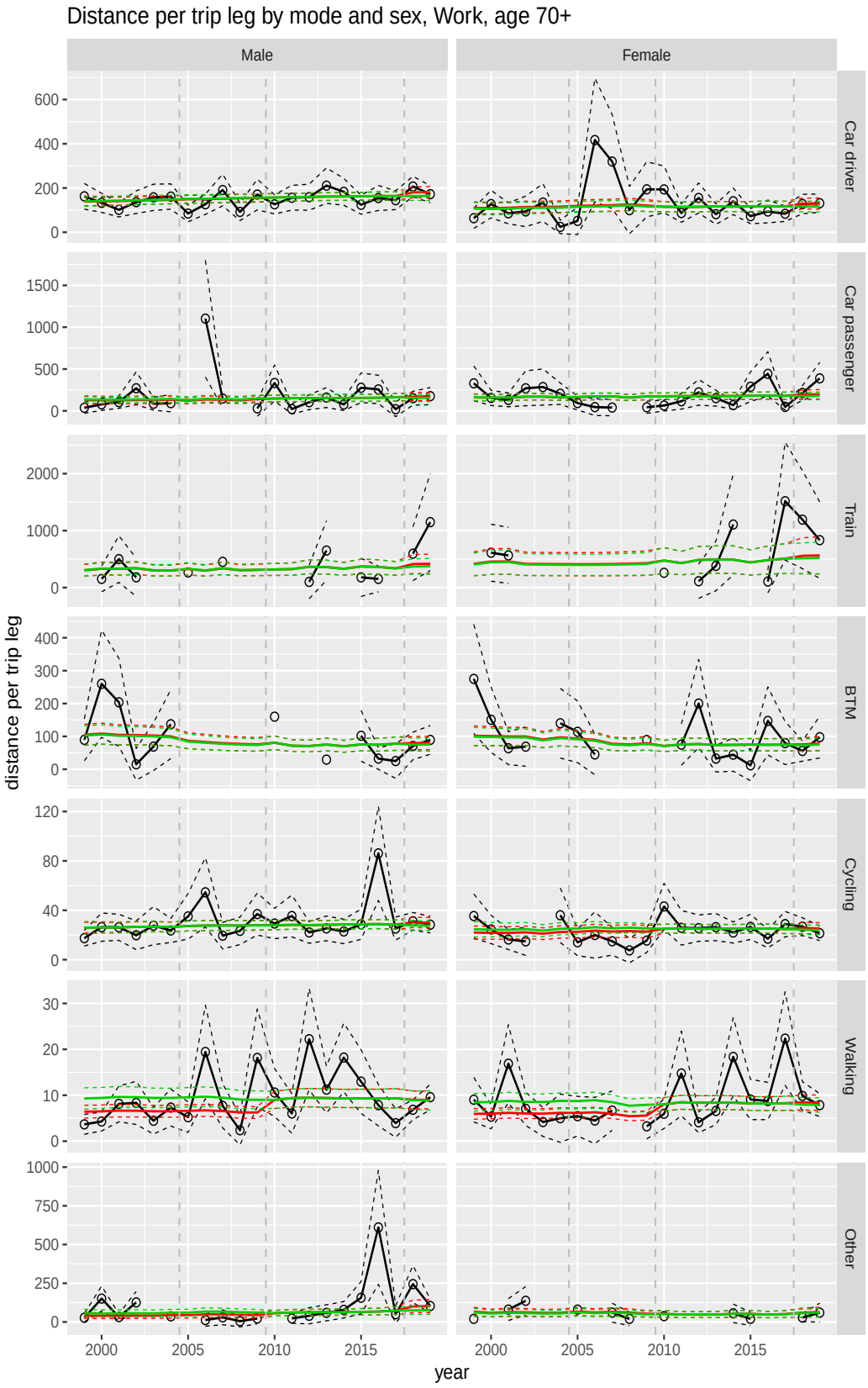

Figure A.105 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

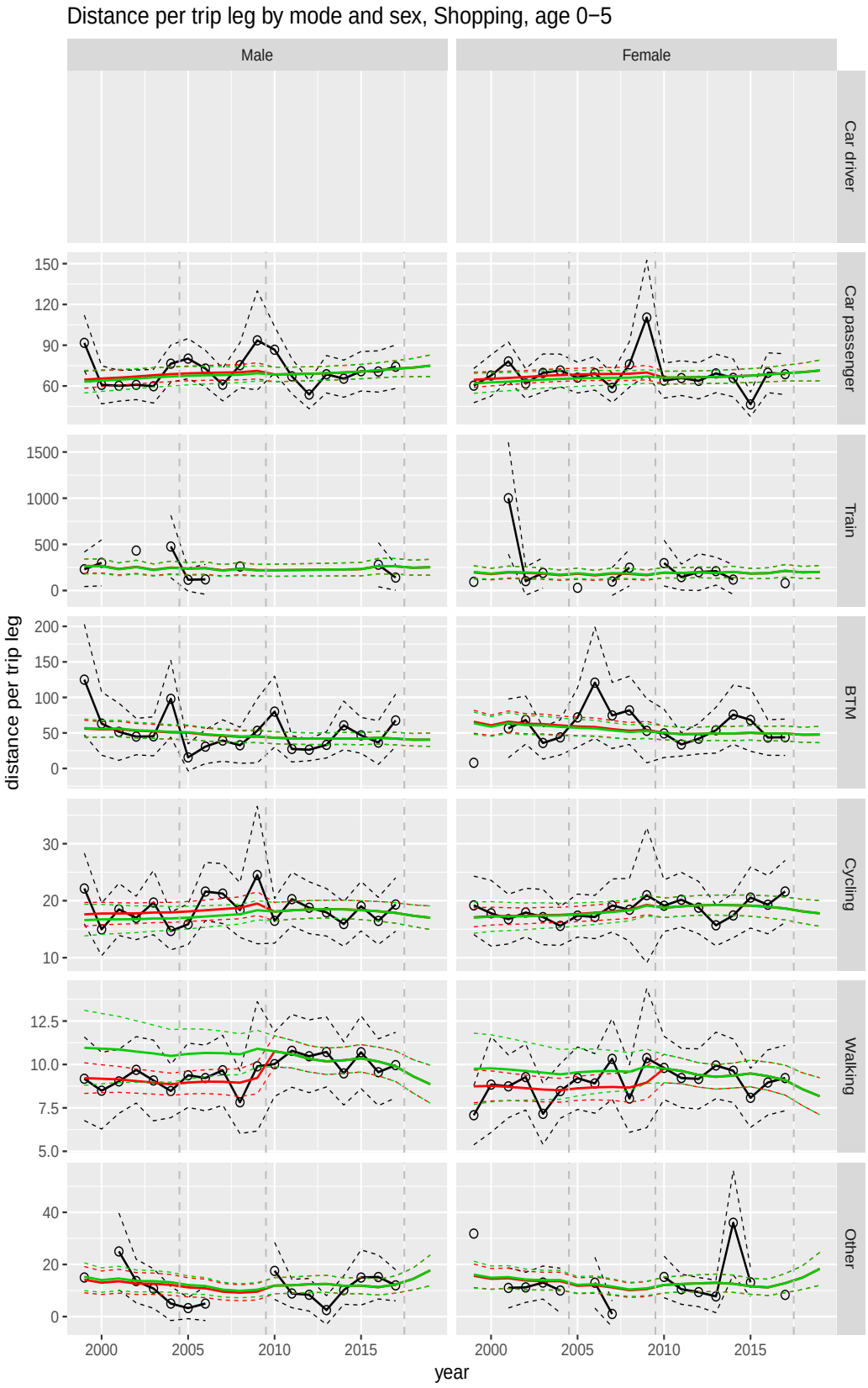

Figure A.106 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

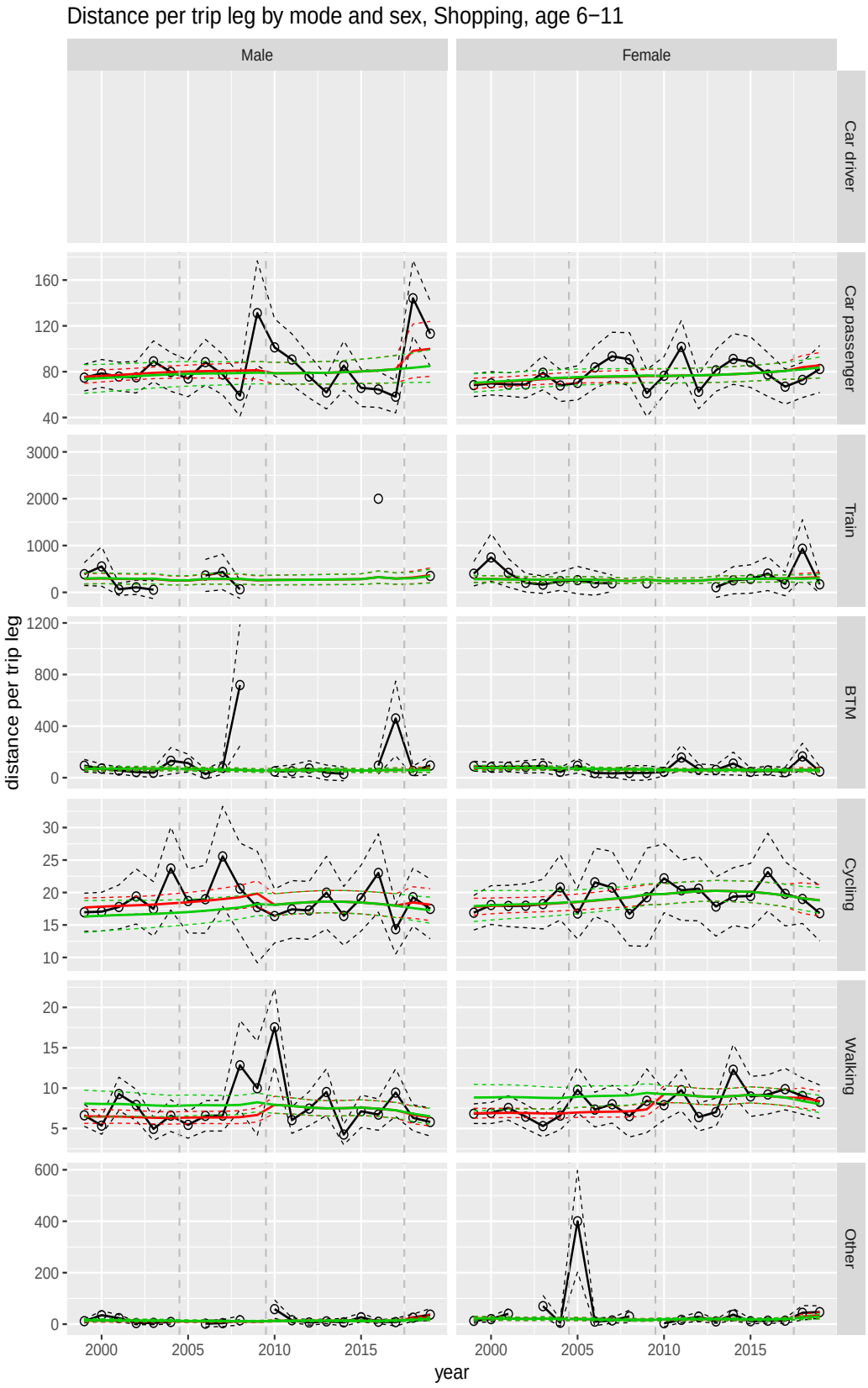

Figure A.107 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

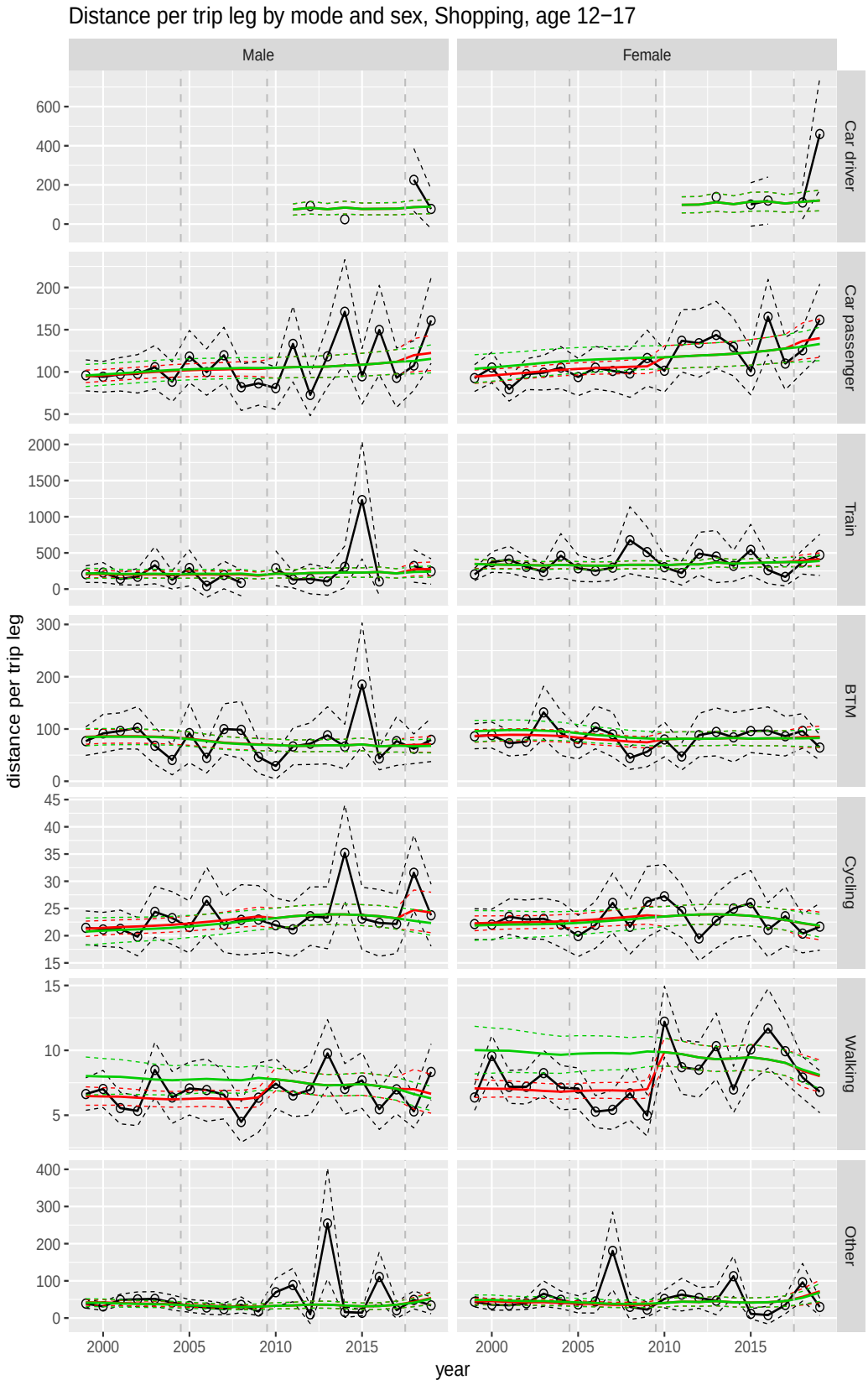

Figure A.108 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

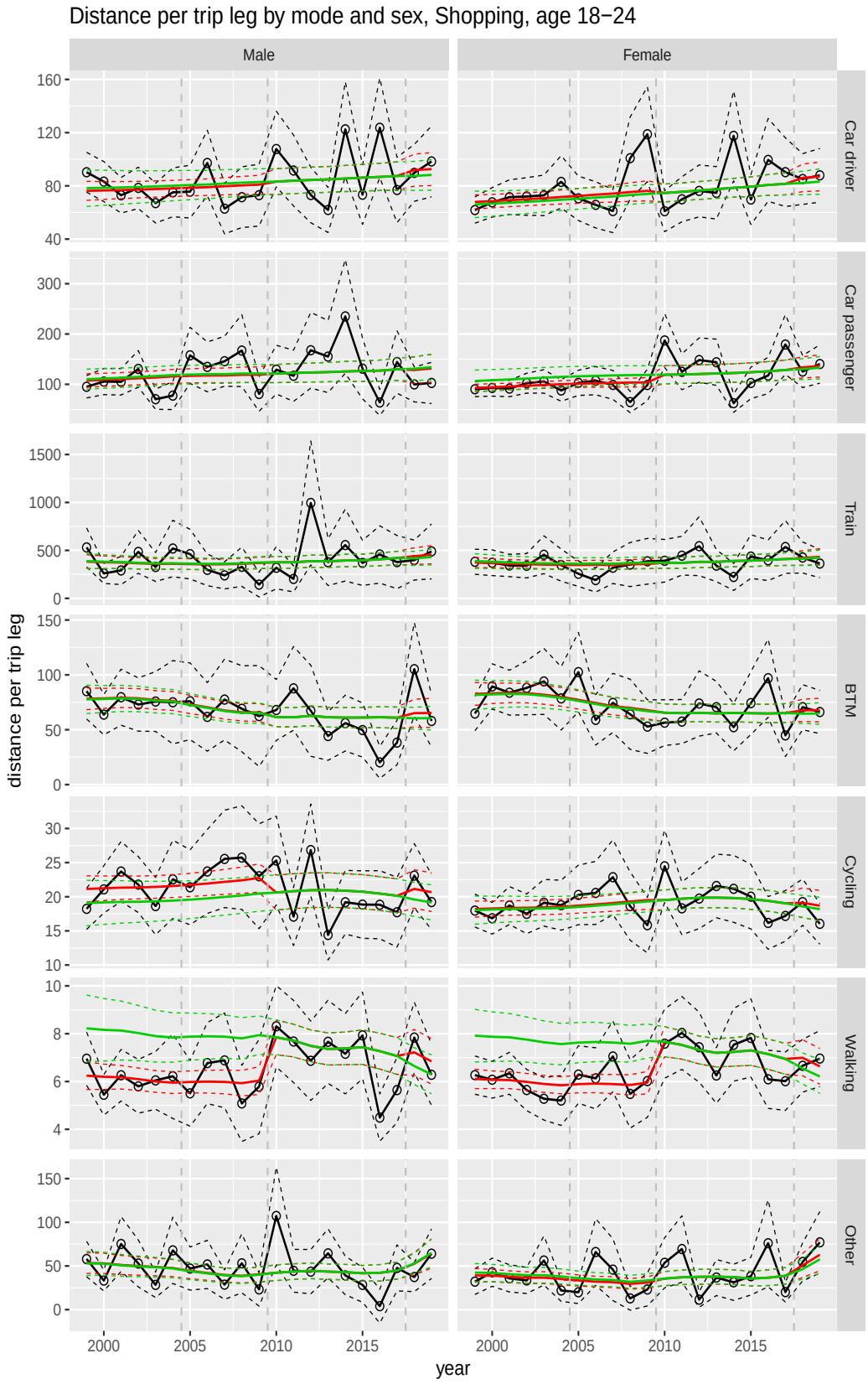

Figure A.109 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

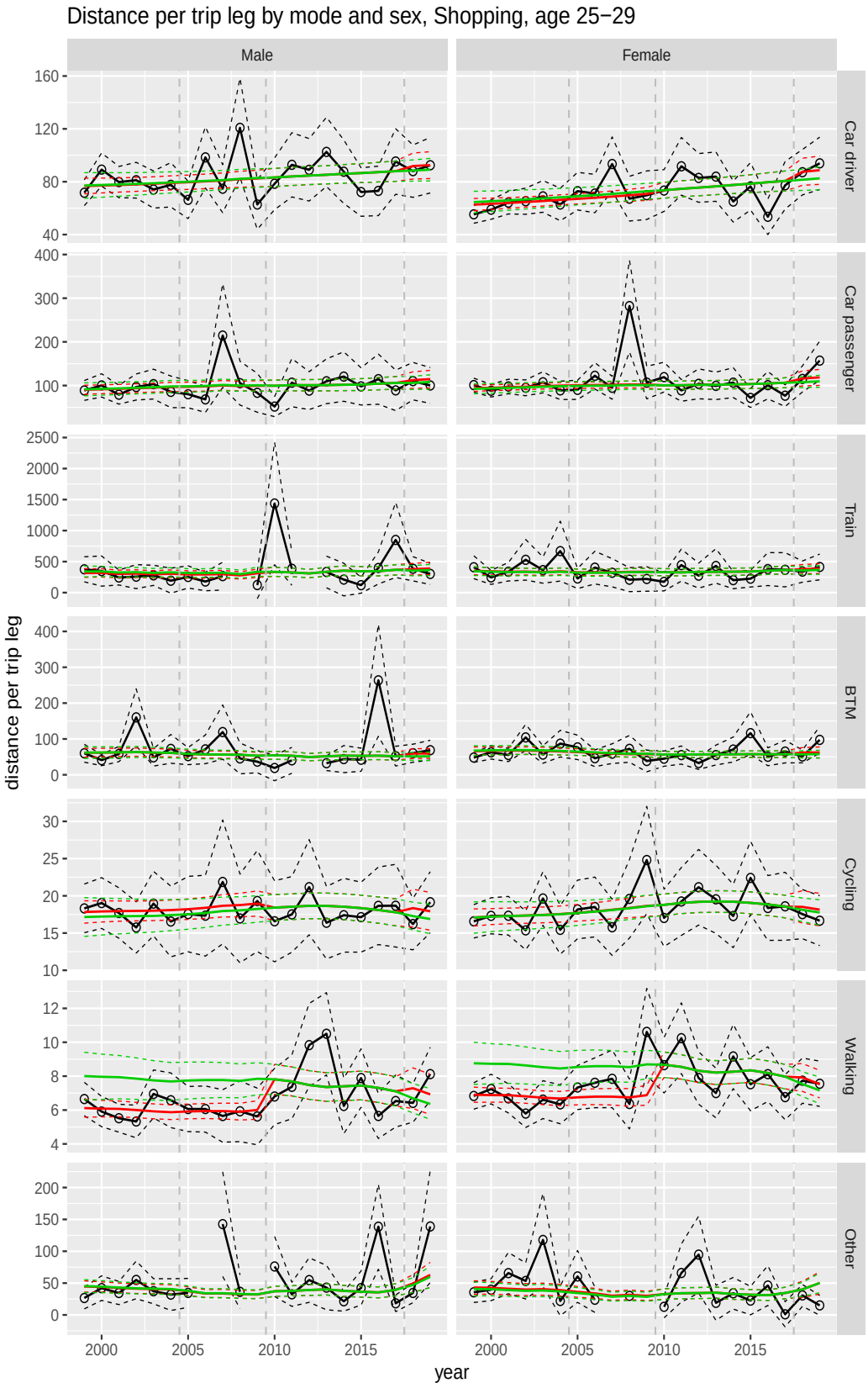

Figure A.110 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

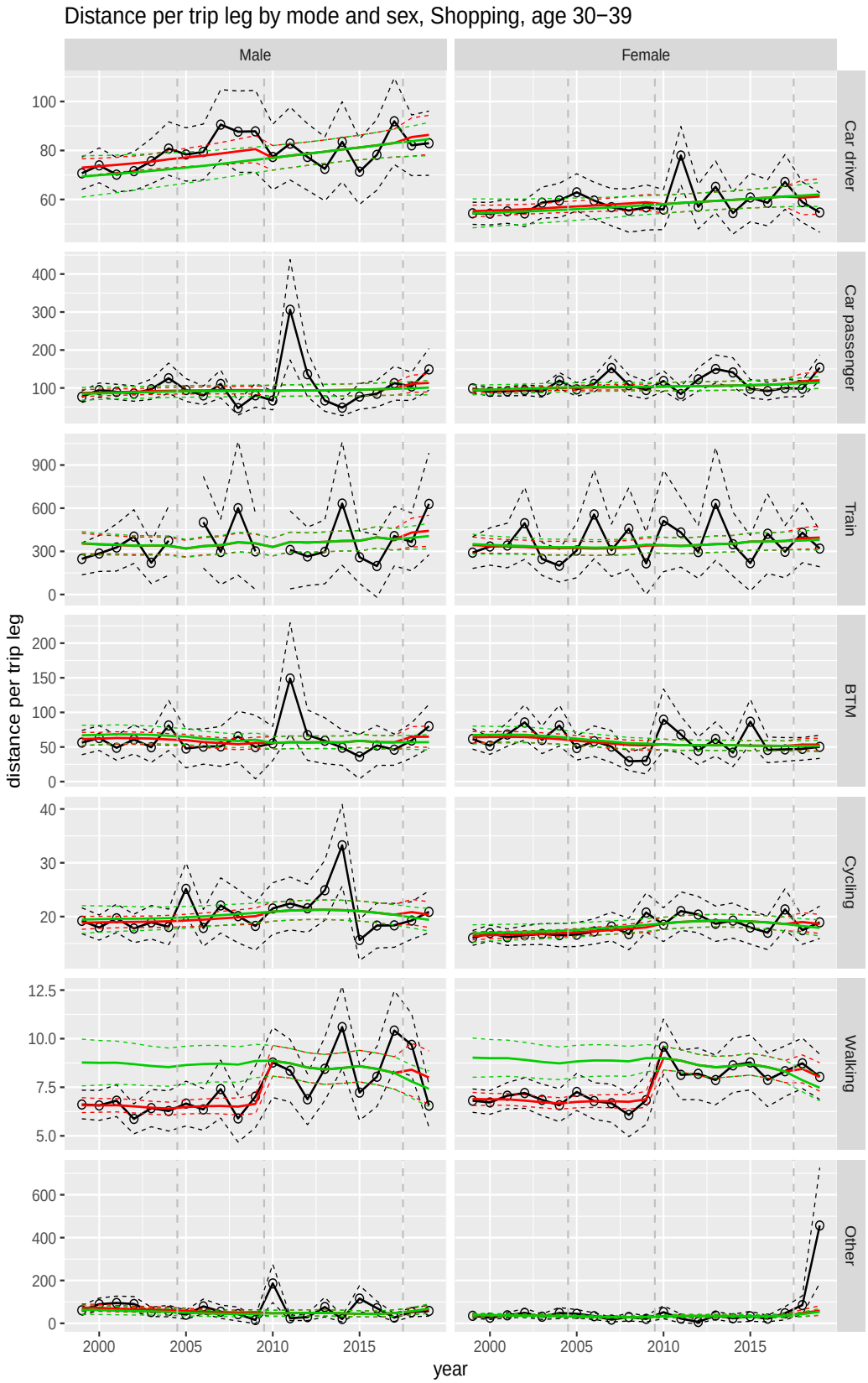

Figure A.111 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

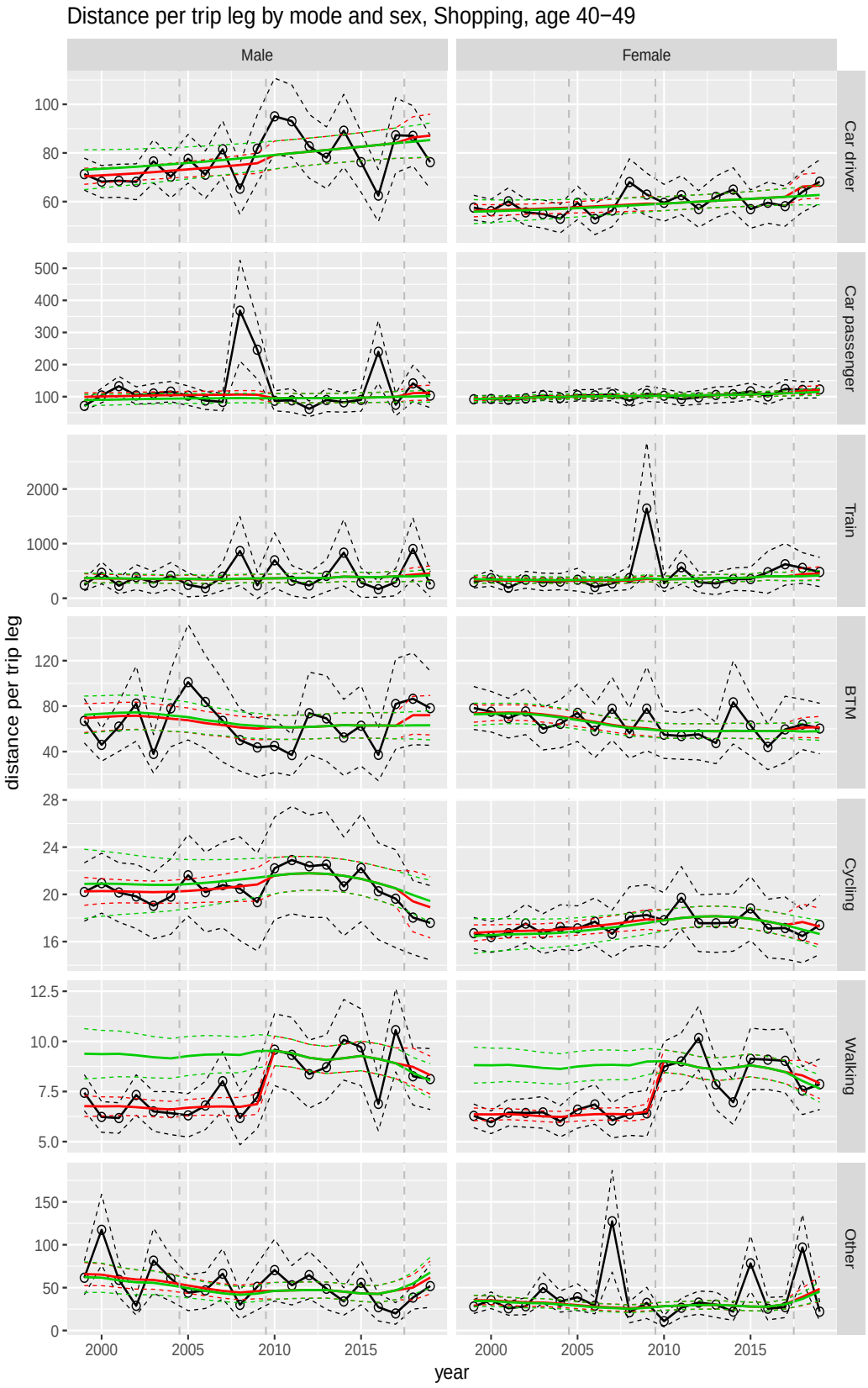

Figure A.112 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

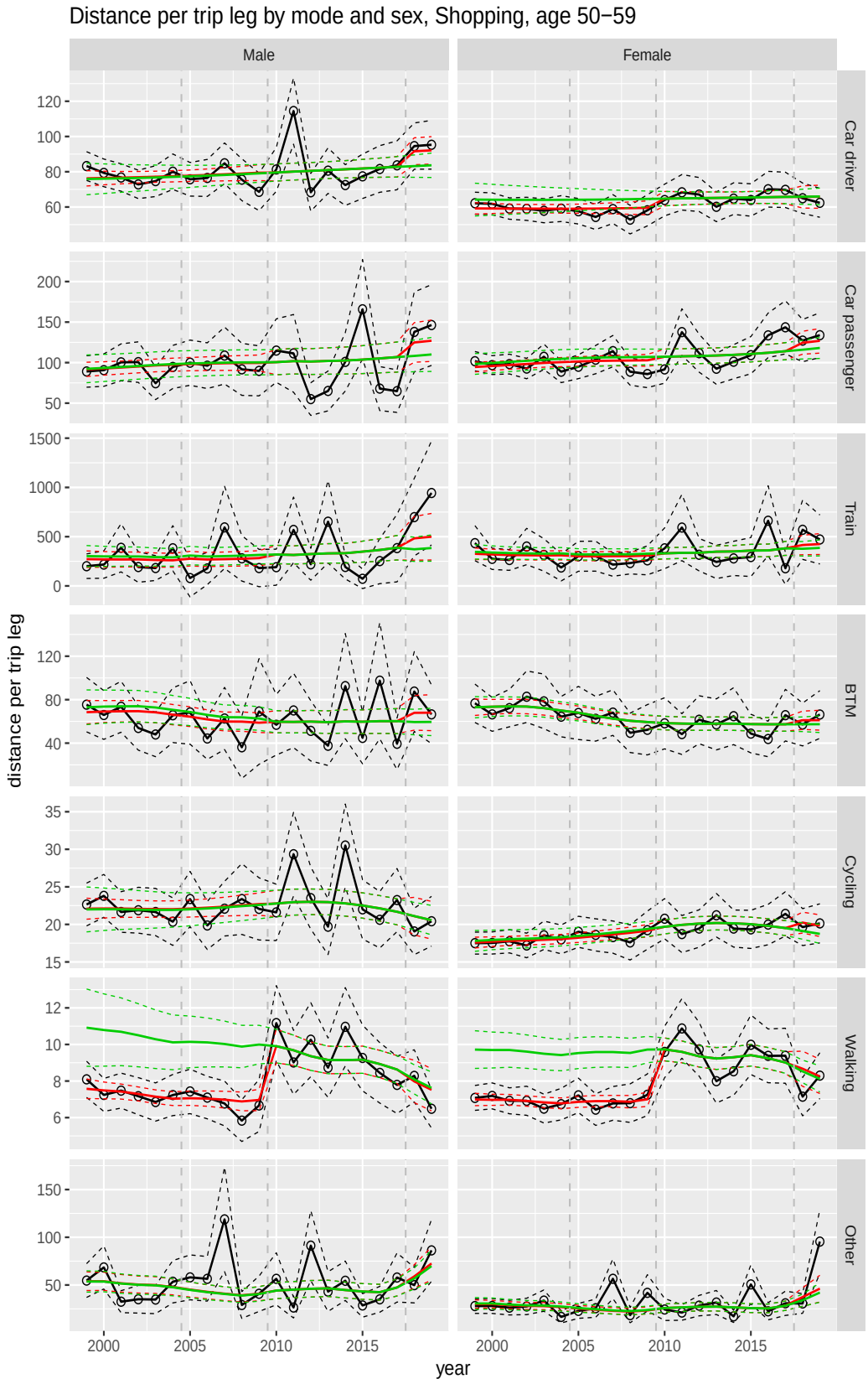

Figure A.113 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

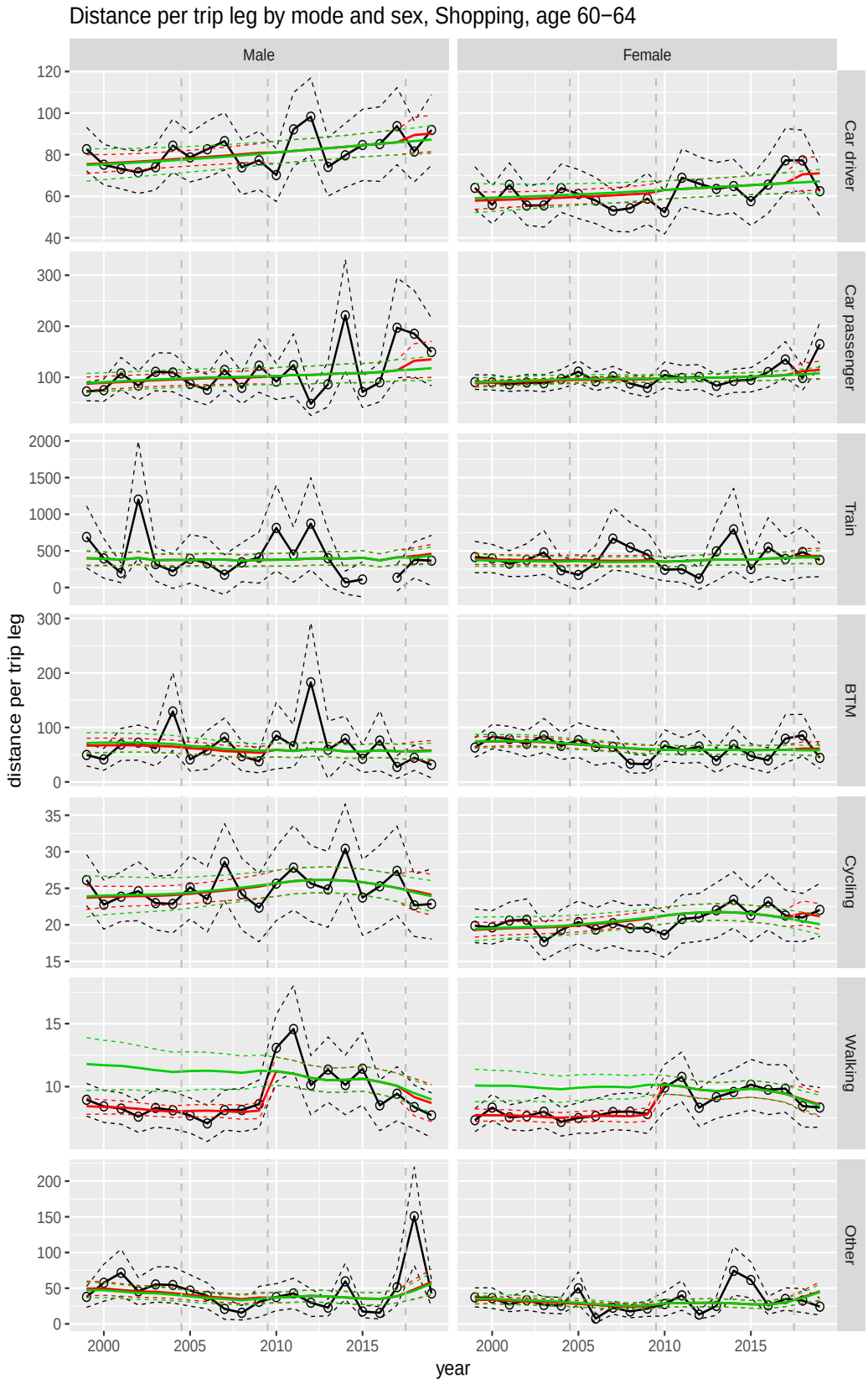

Figure A.114 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

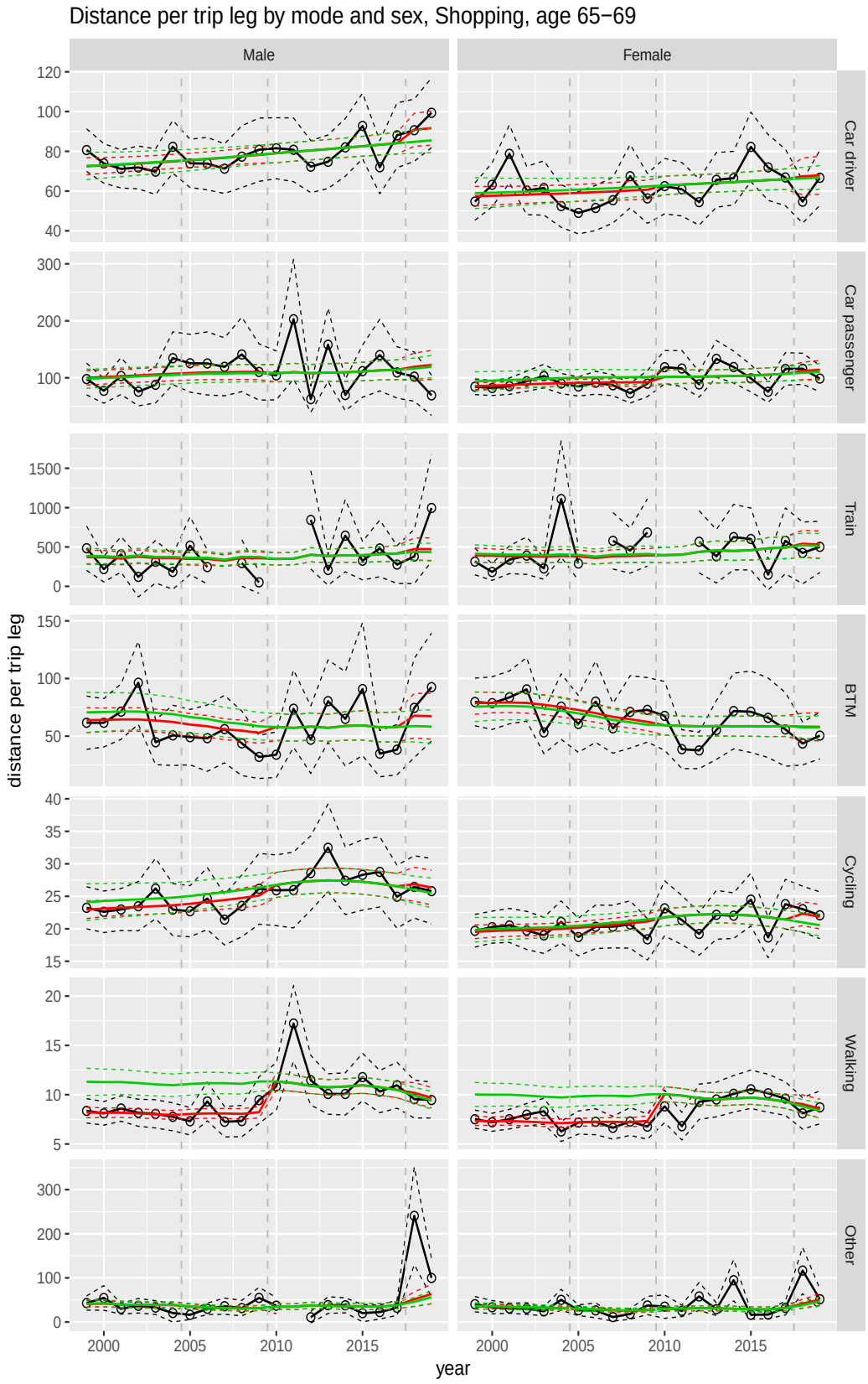

Figure A.115 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

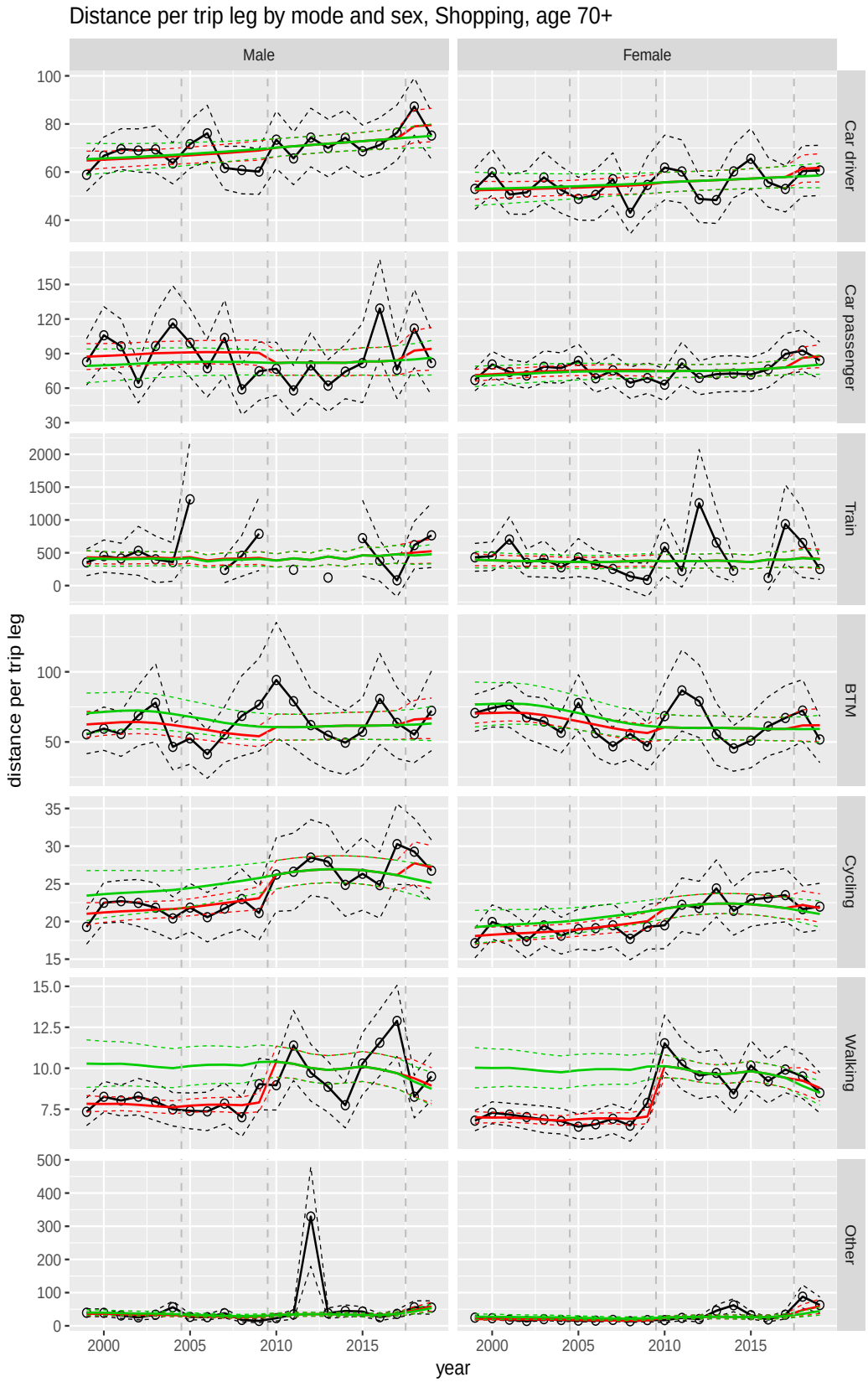

Figure A.116 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

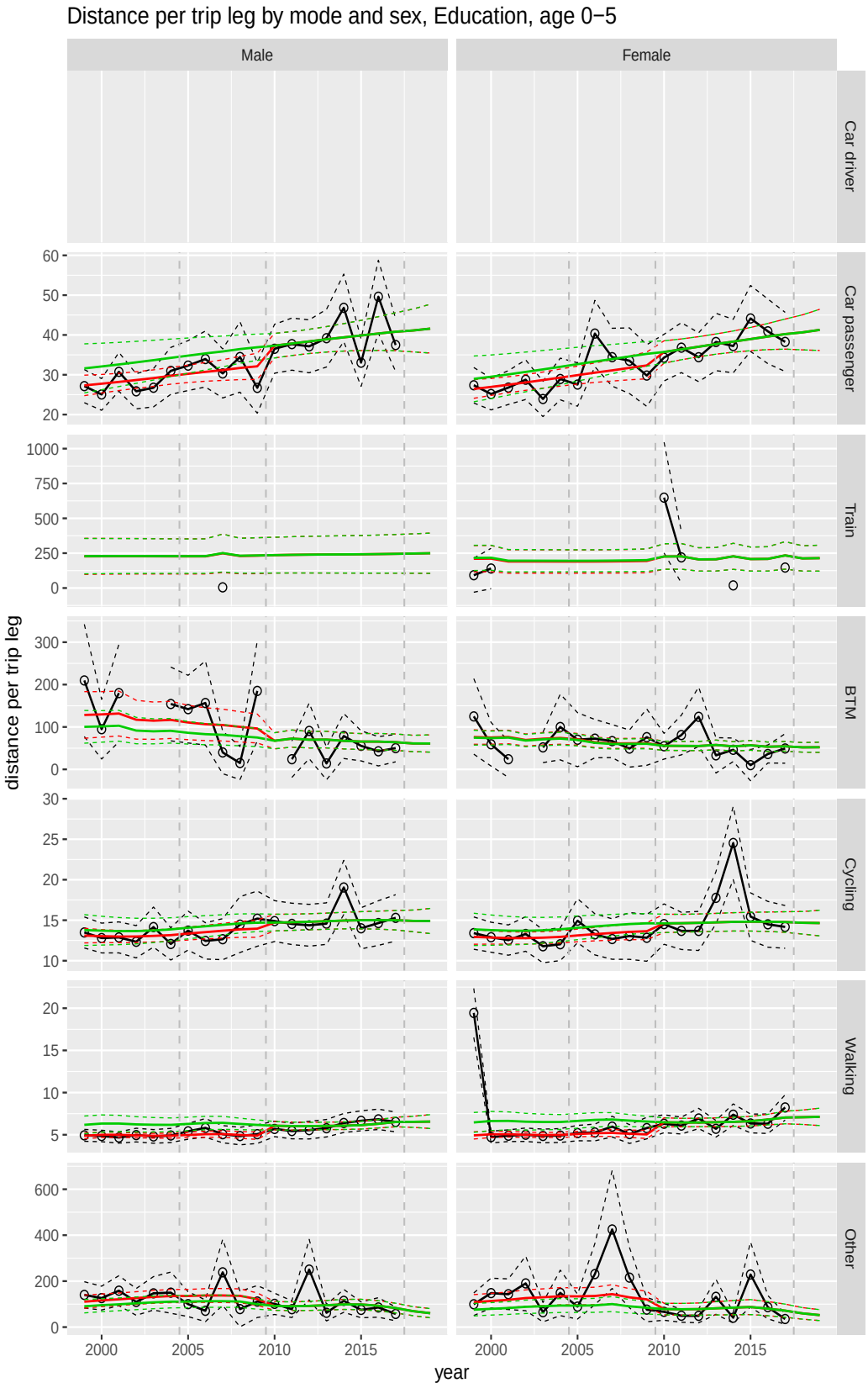

Figure A.117 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

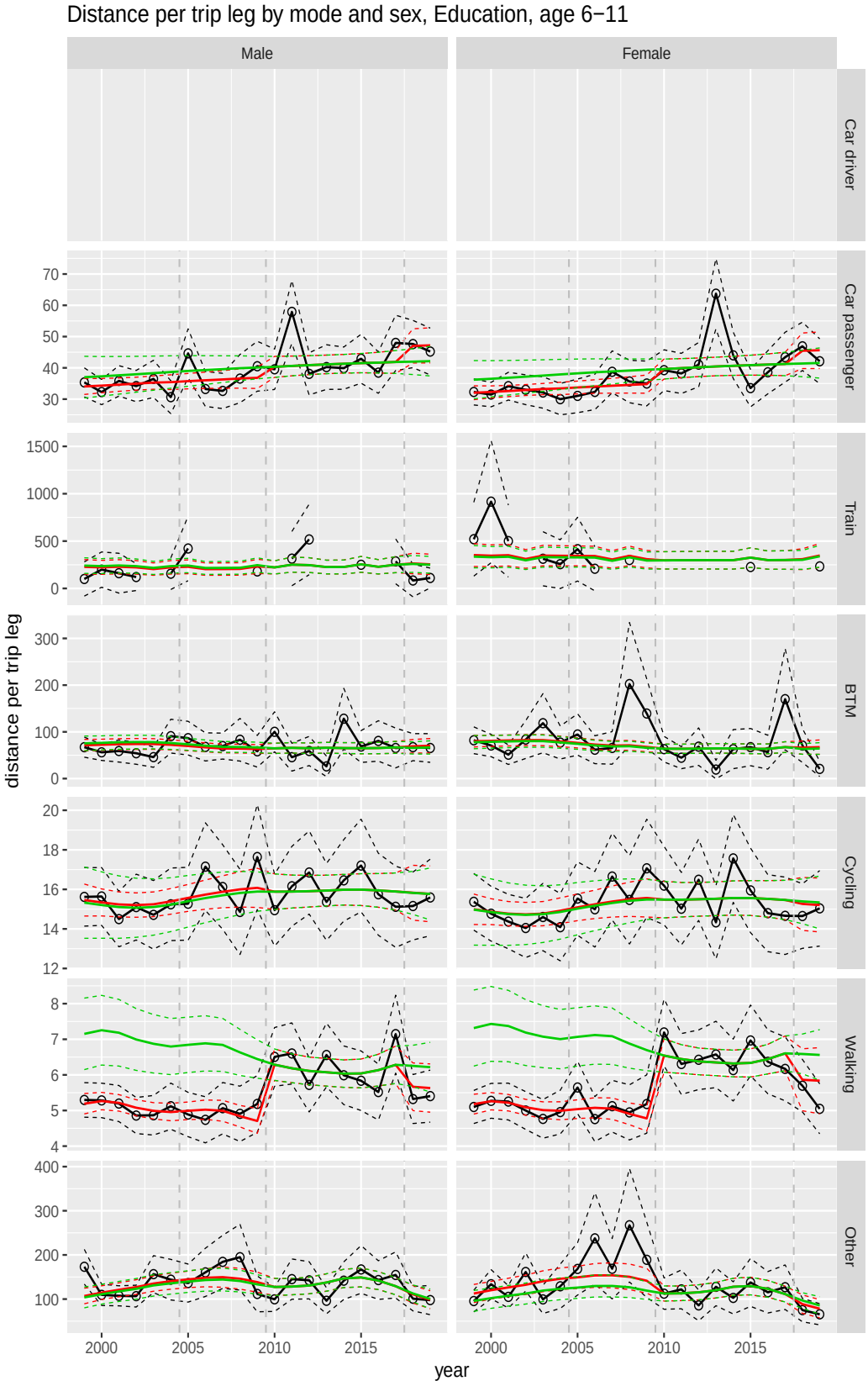

Figure A.118 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

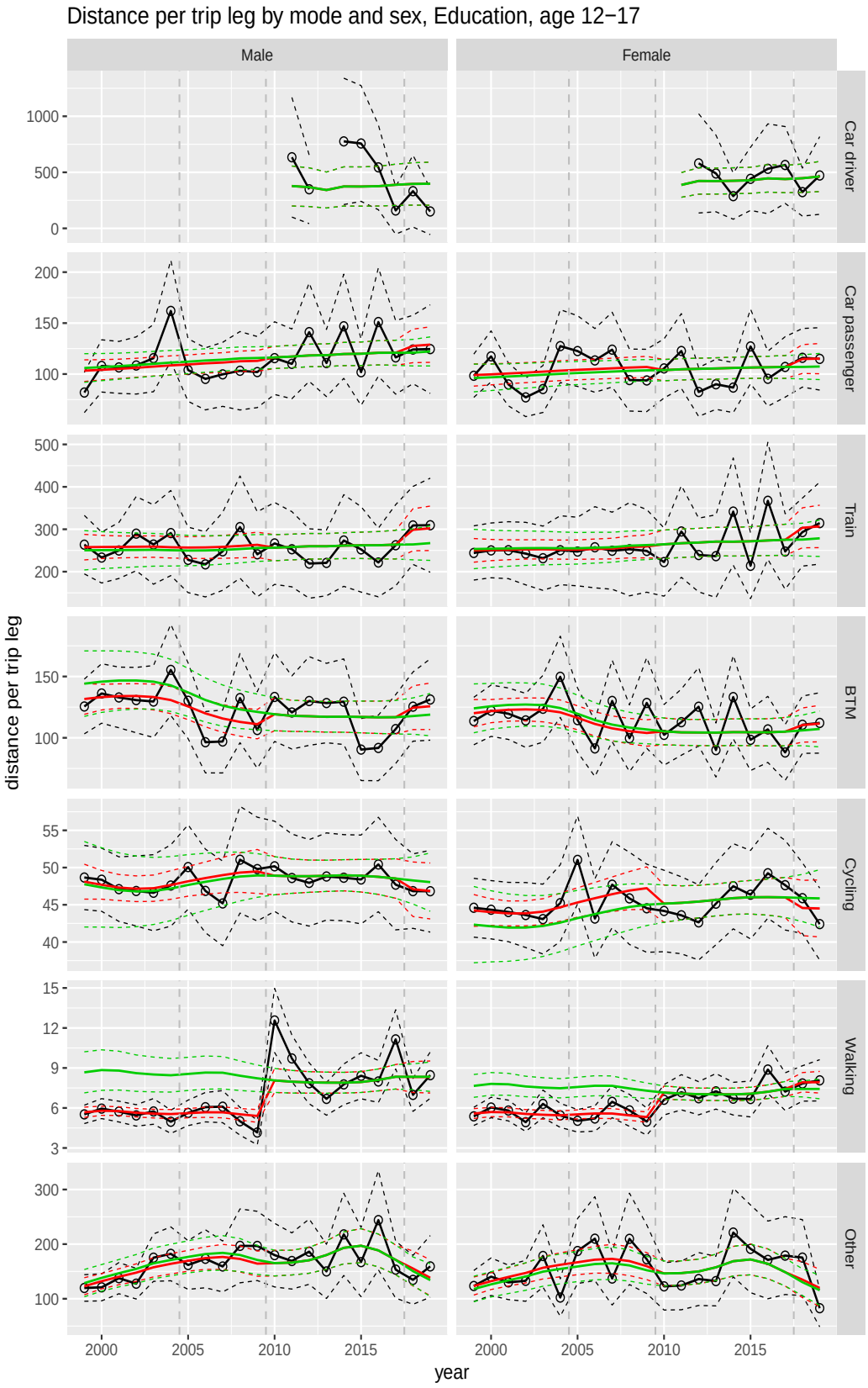

Figure A.119 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

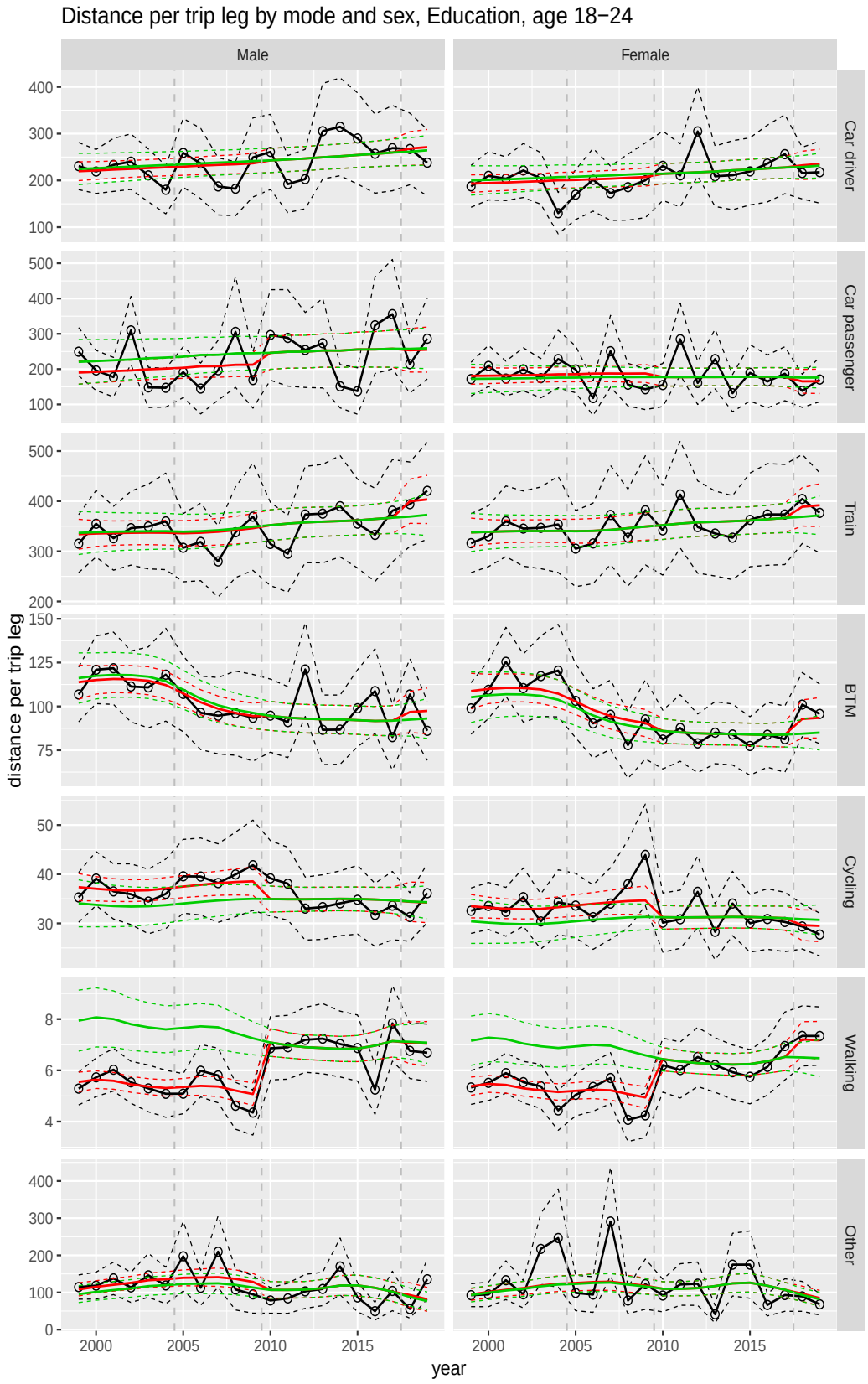

Figure A.120 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

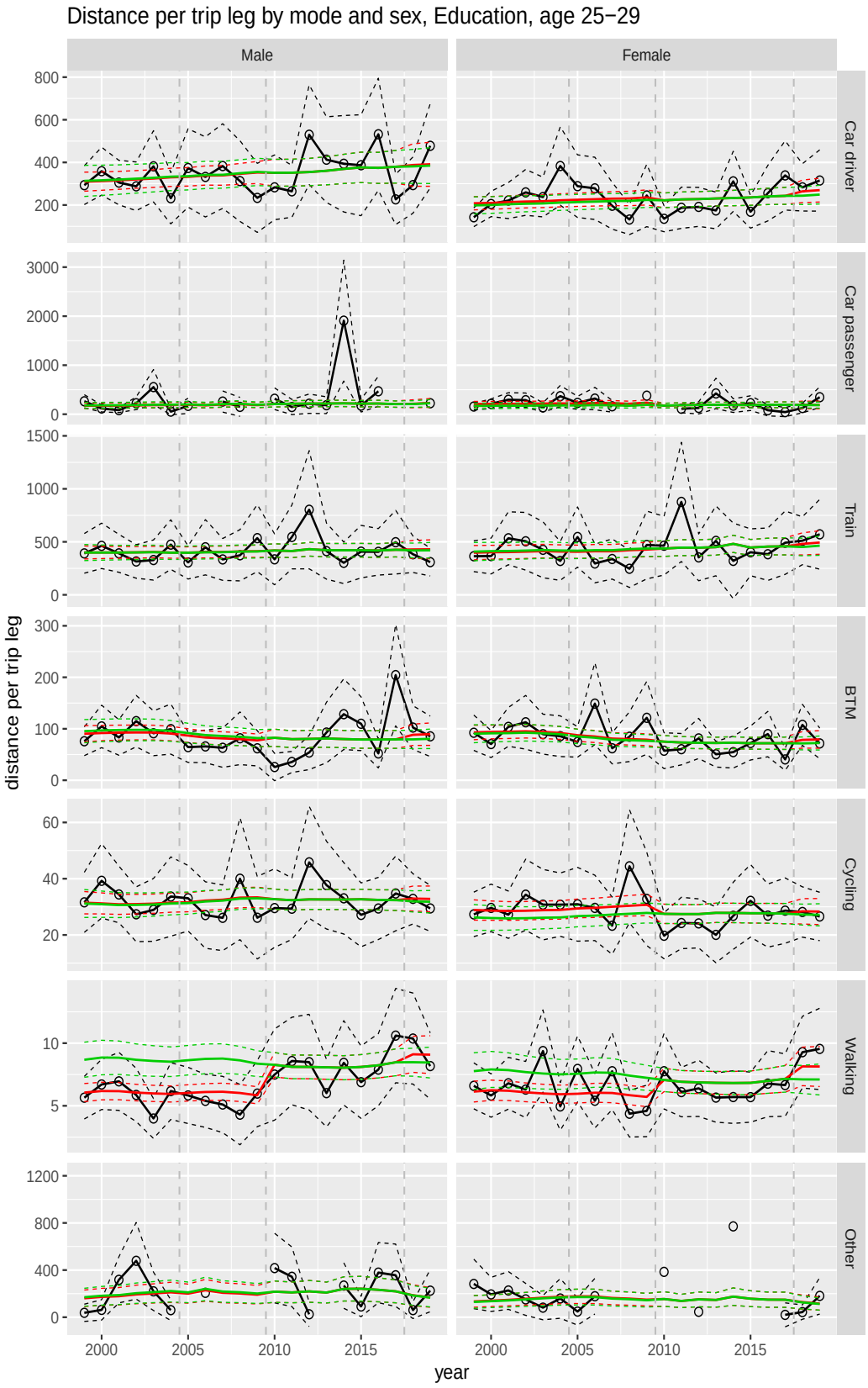

Figure A.121 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

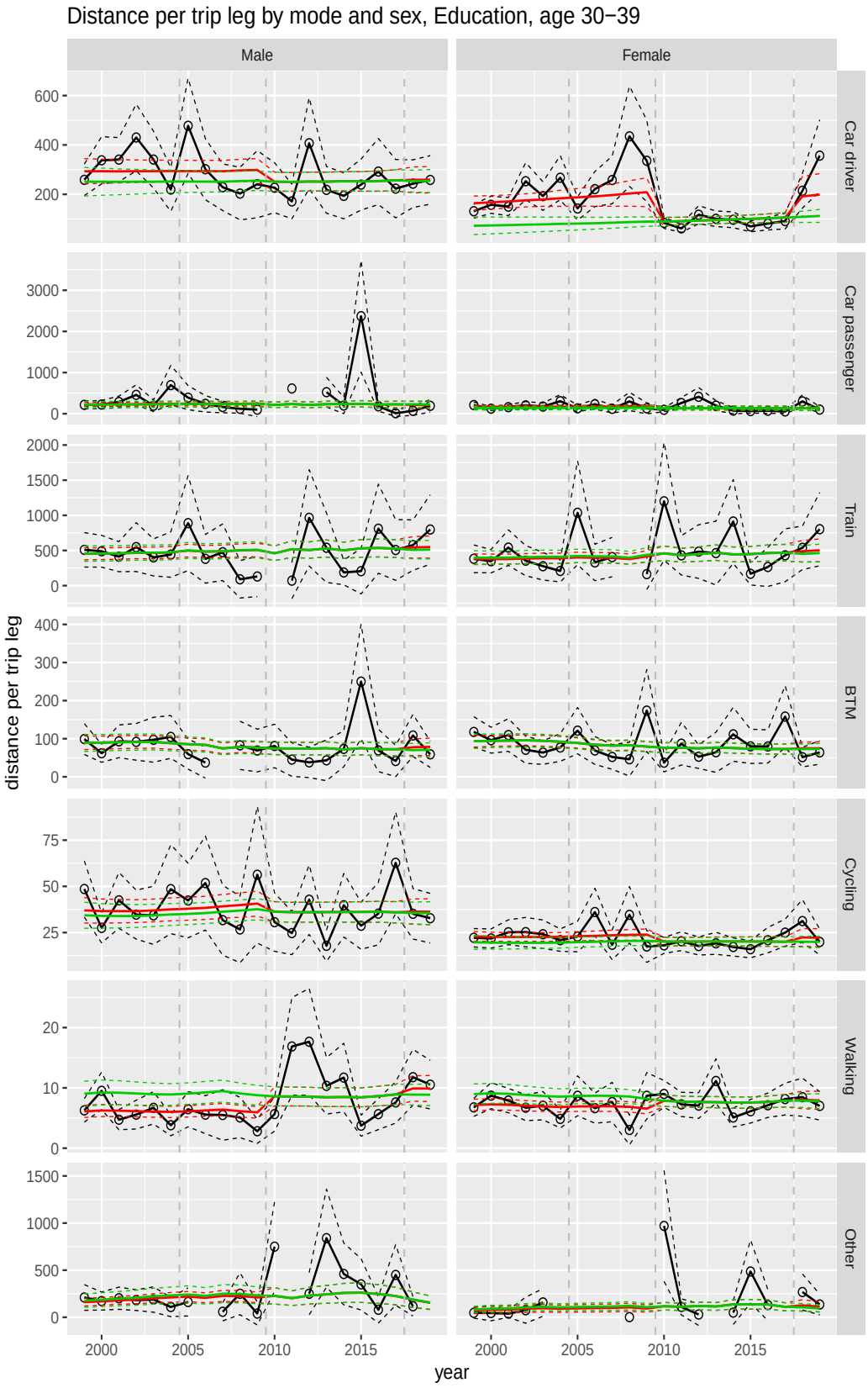

Figure A.122 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

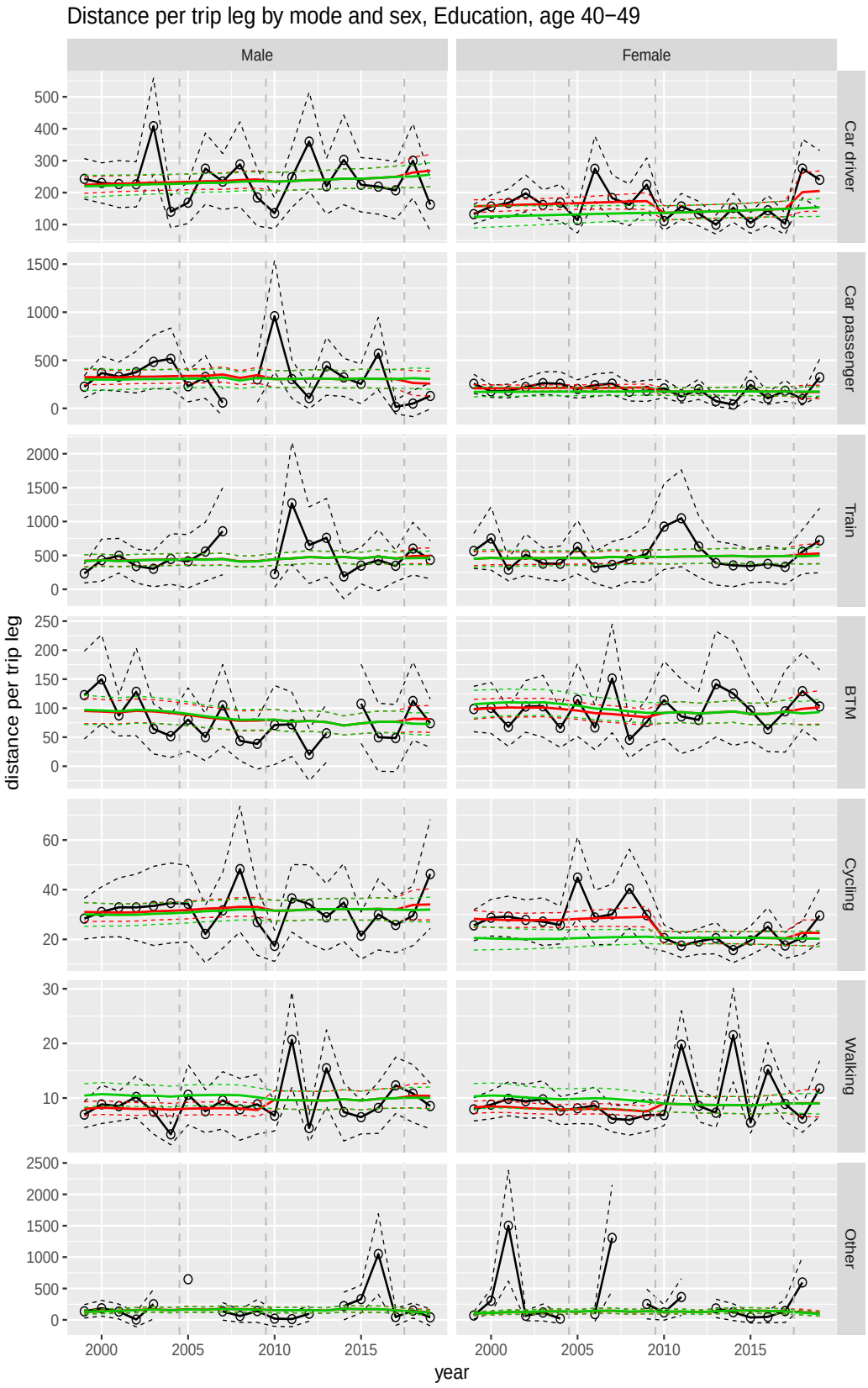

Figure A.123 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

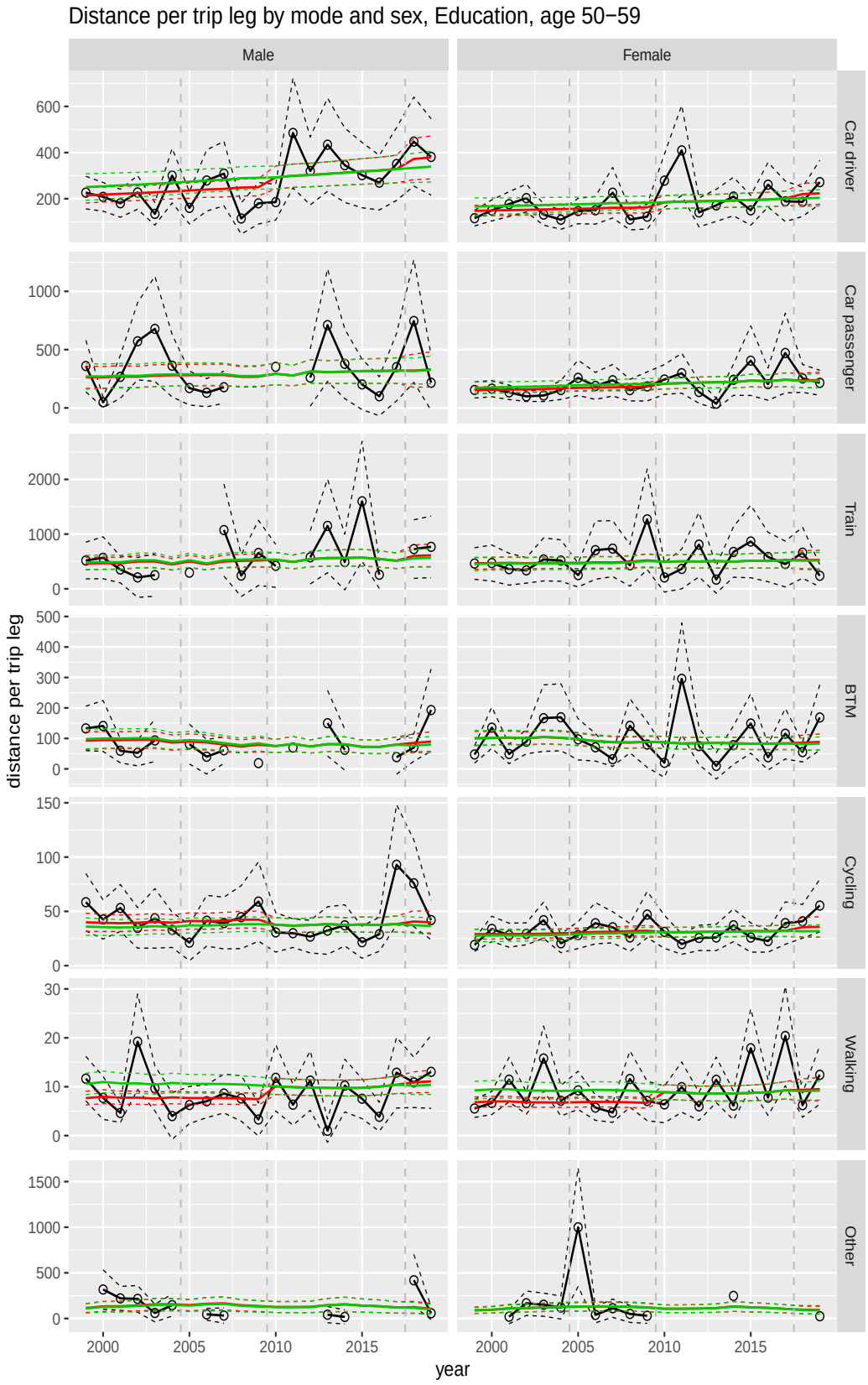

Figure A.124 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

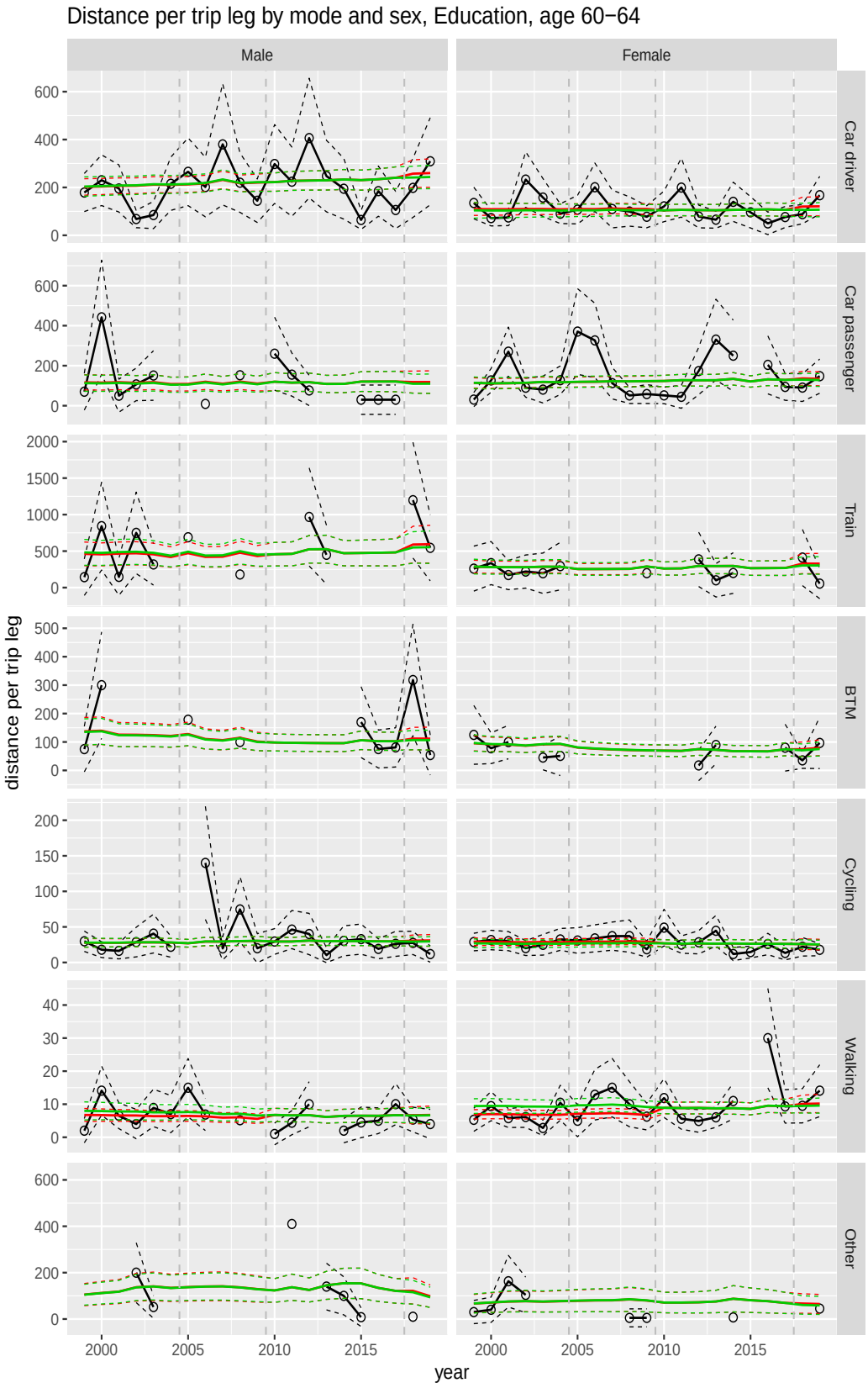

Figure A.125 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

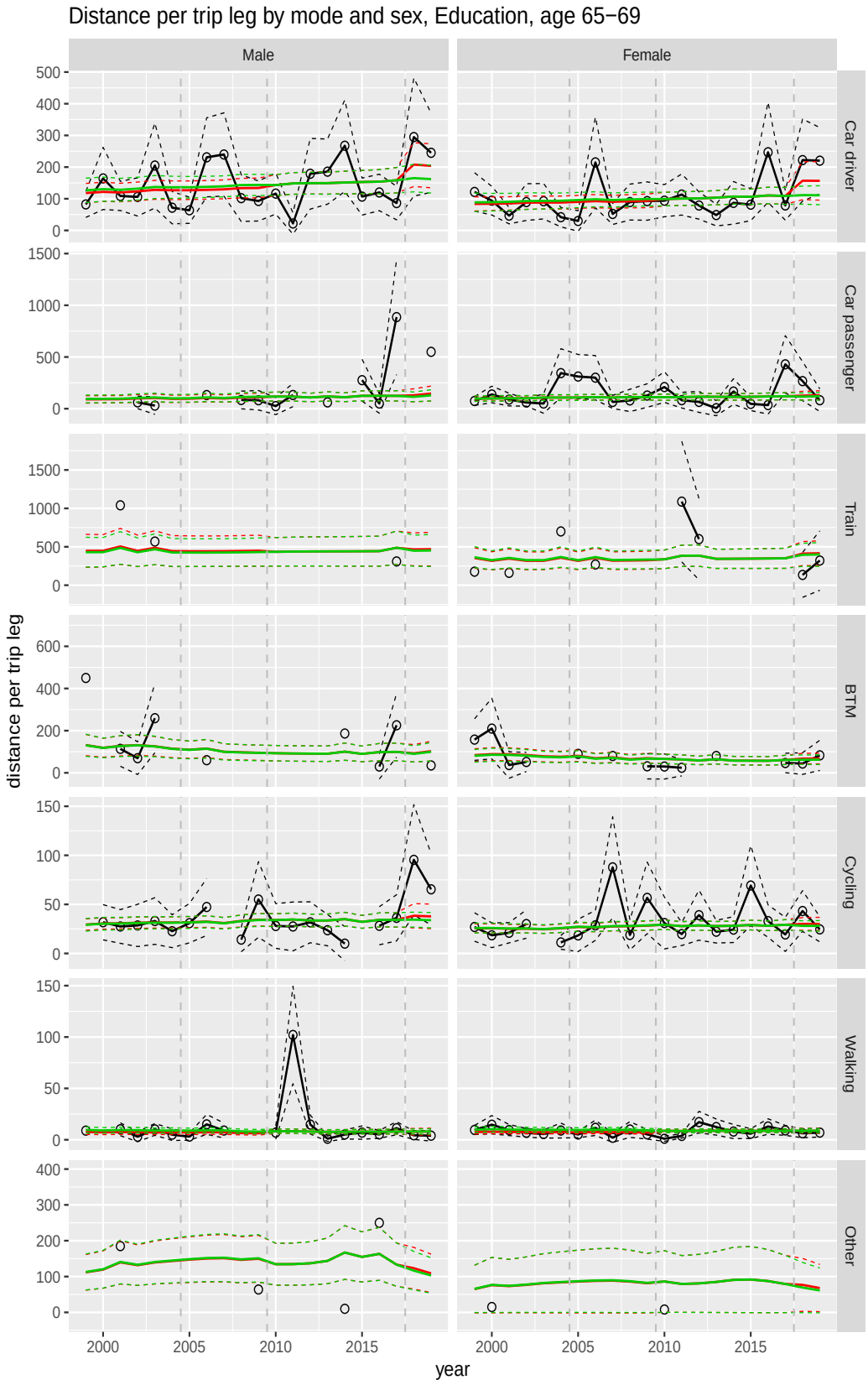

Figure A.126 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

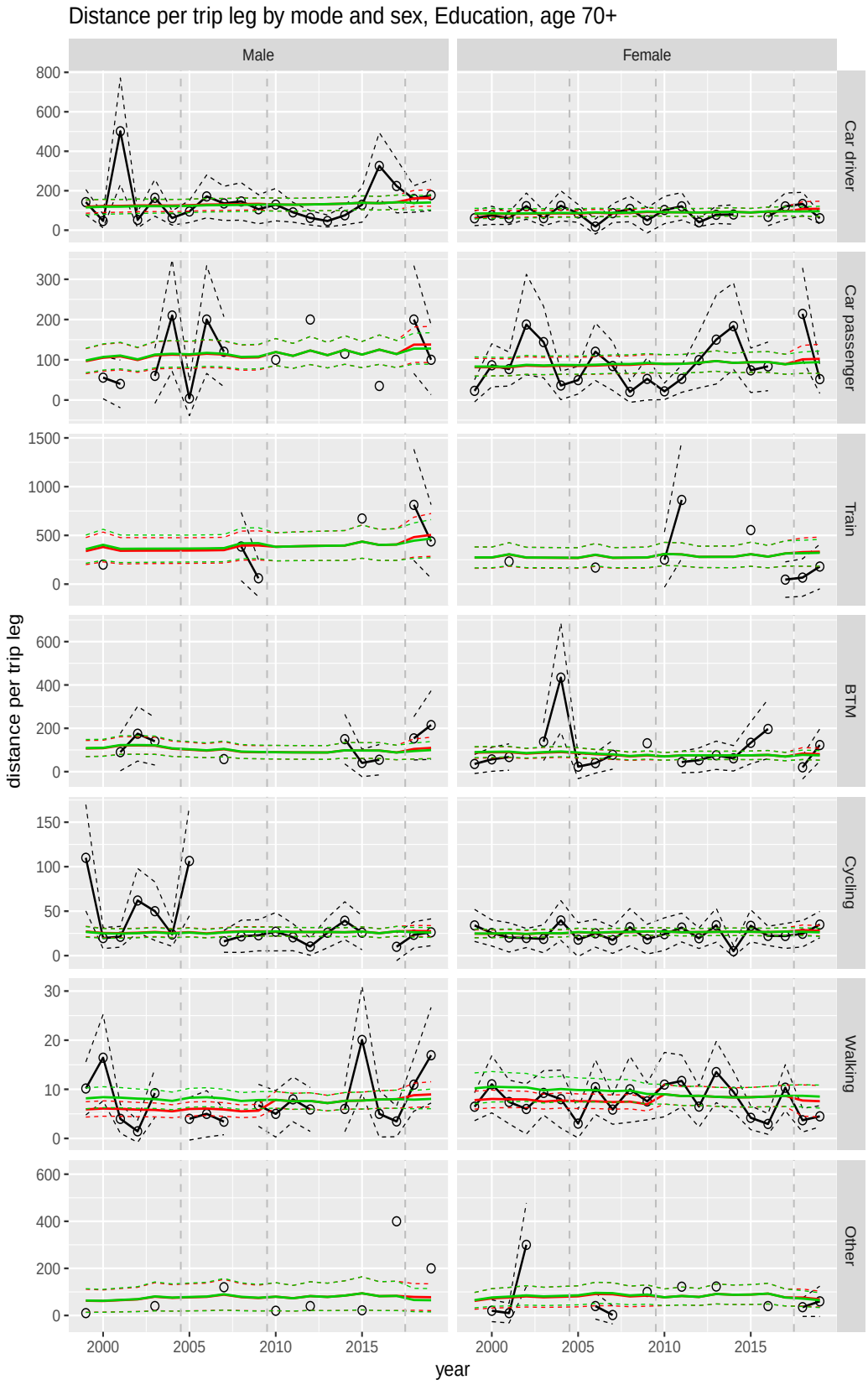

Figure A.127 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

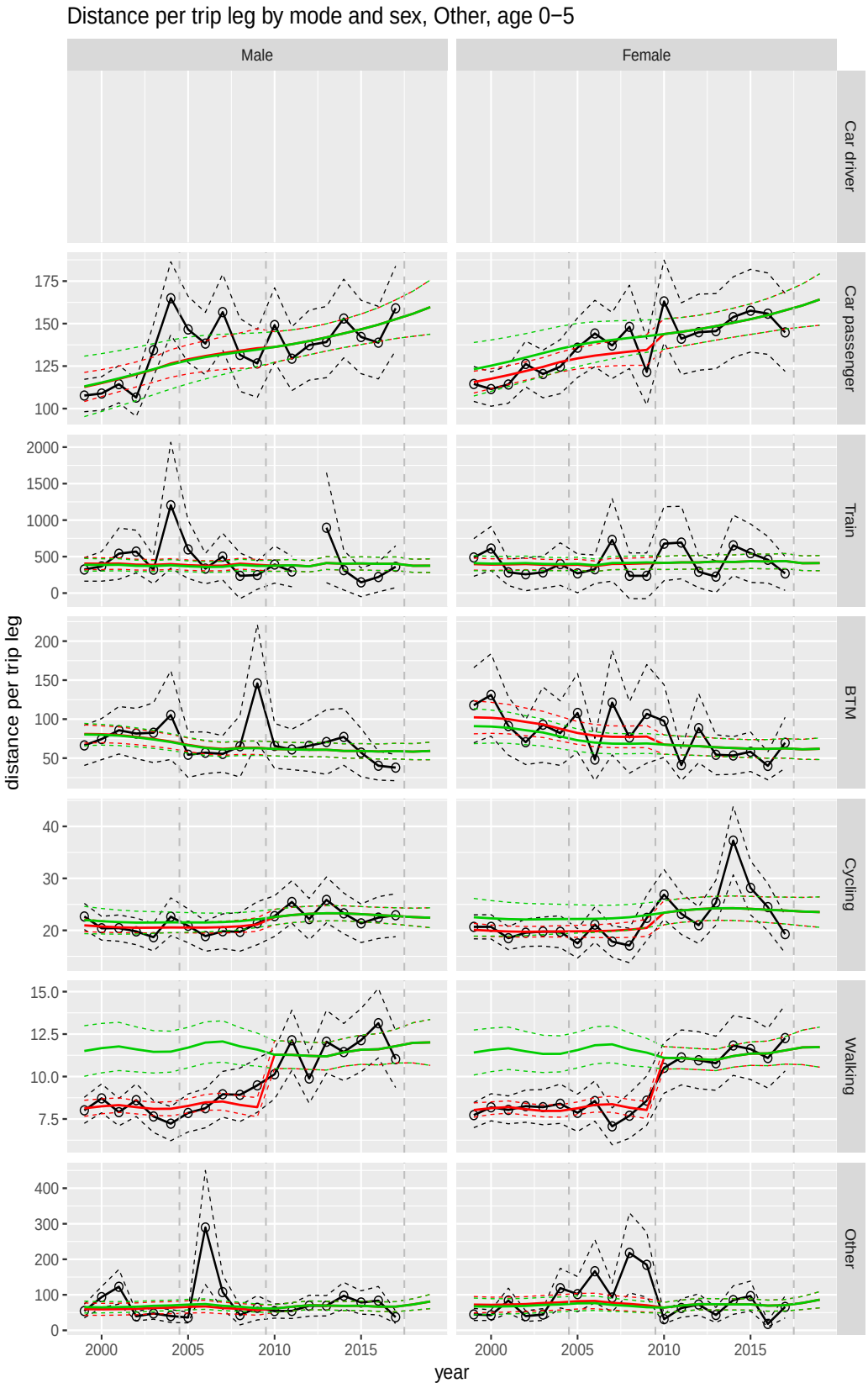

Figure A.128 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

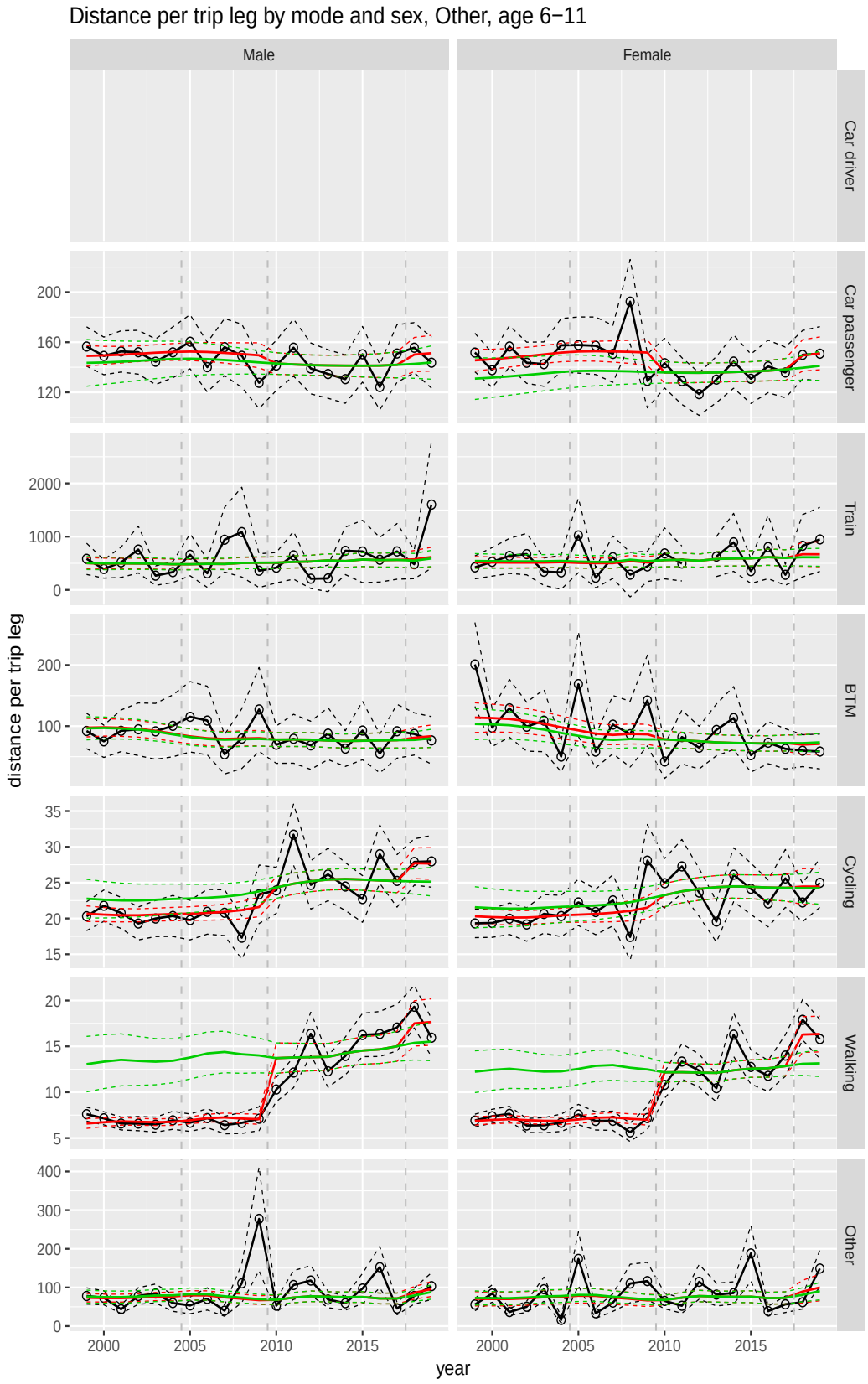

Figure A.129 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

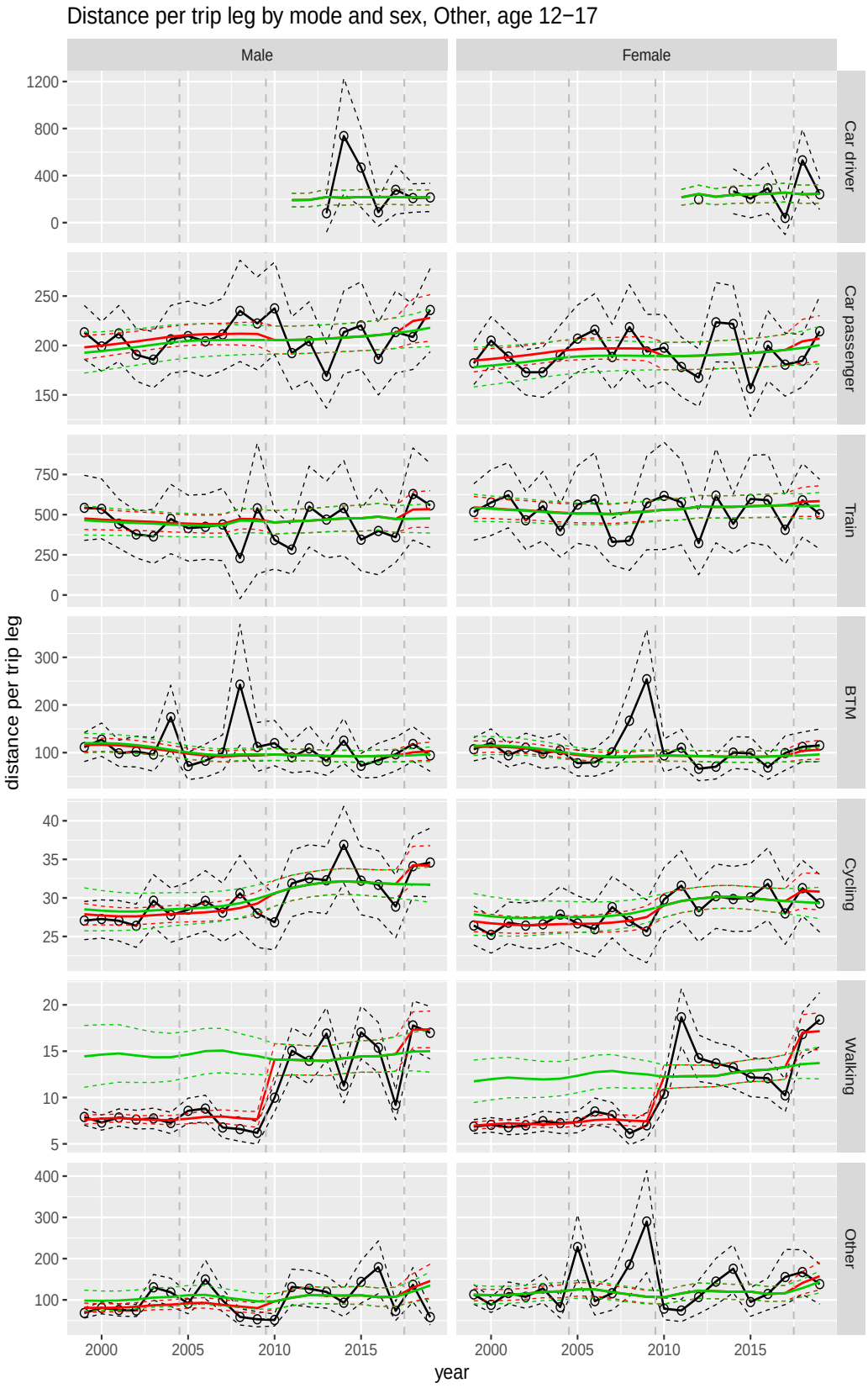

Figure A.130 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

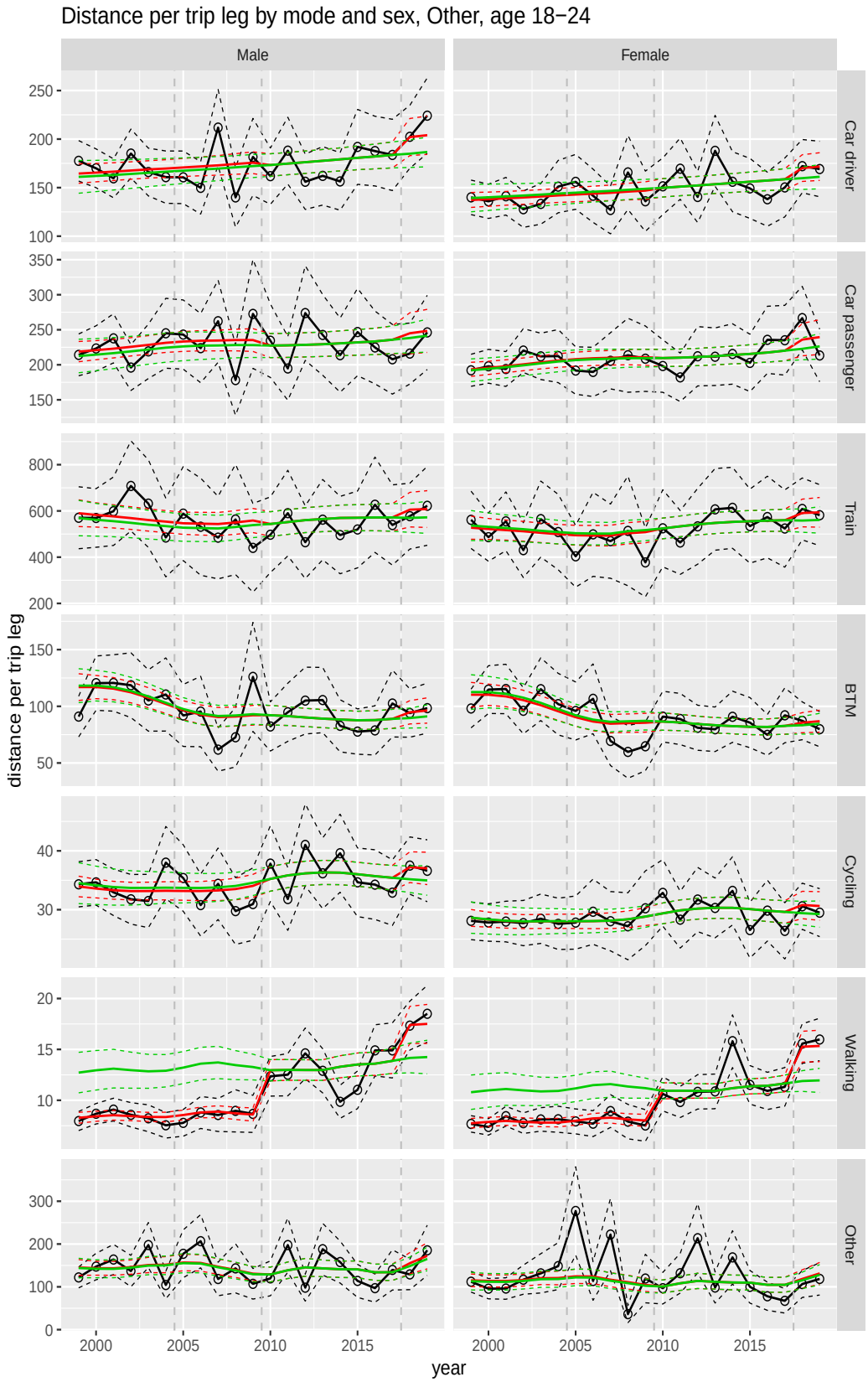

Figure A.131 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

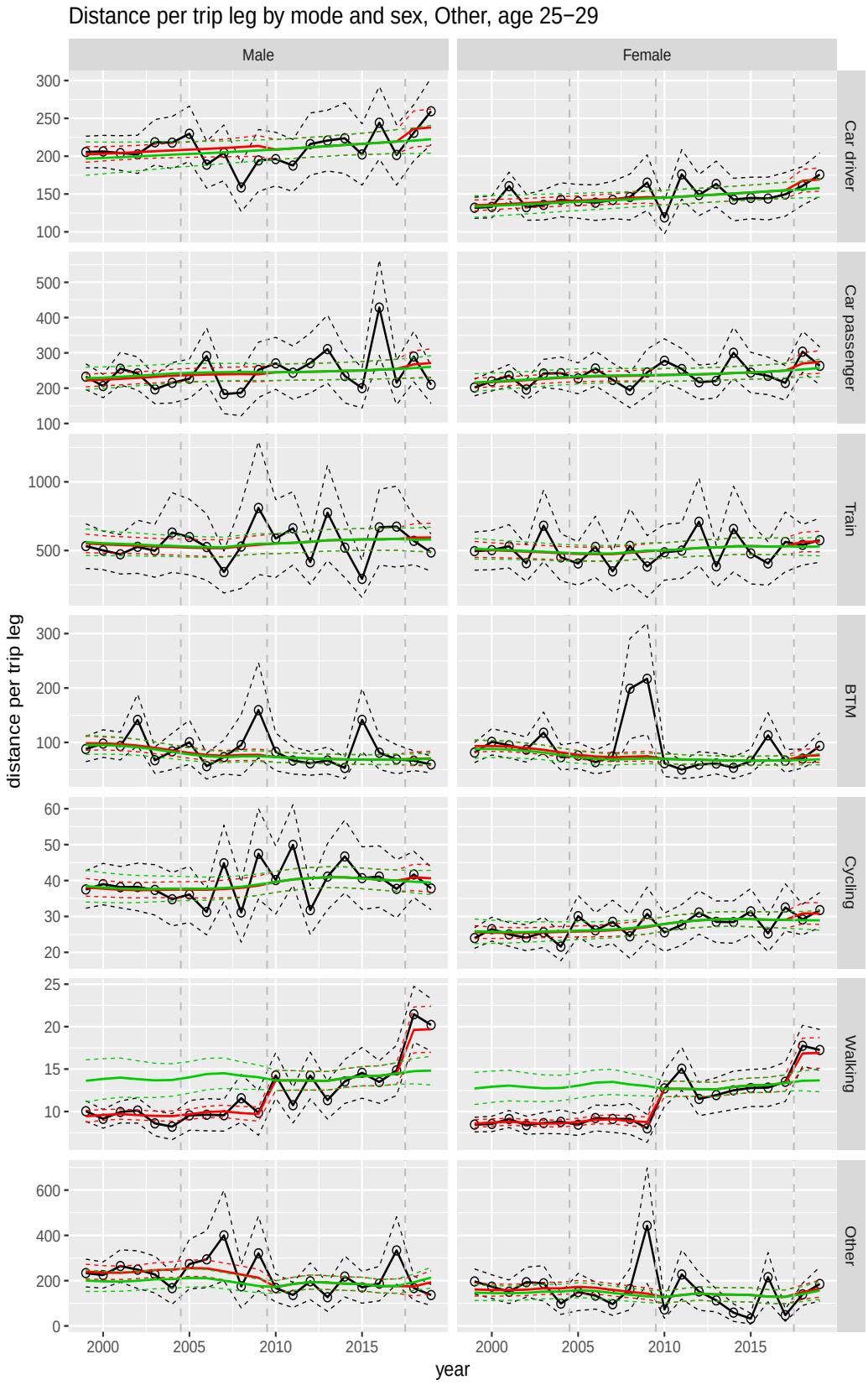

Figure A.132 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

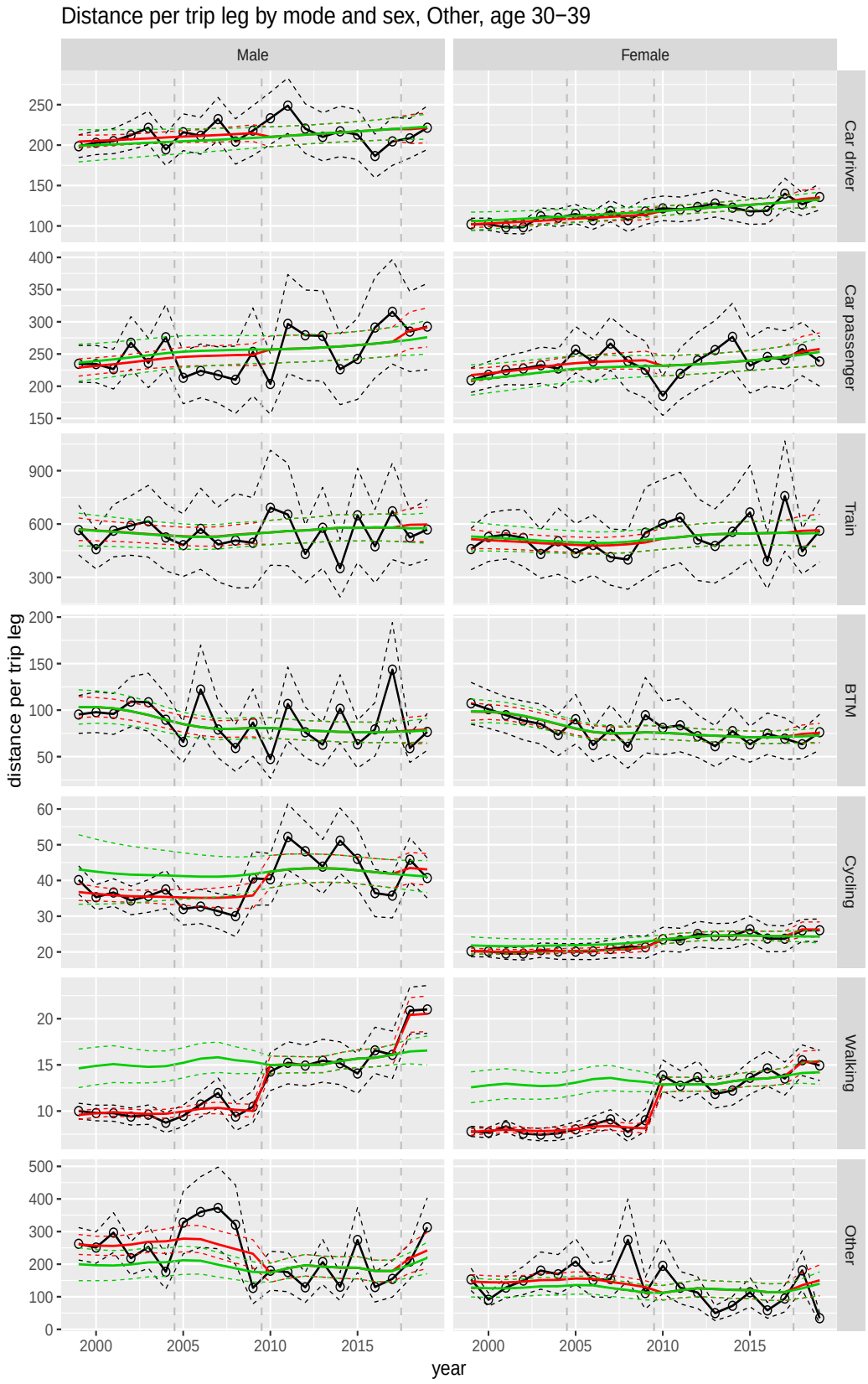

Figure A.133 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

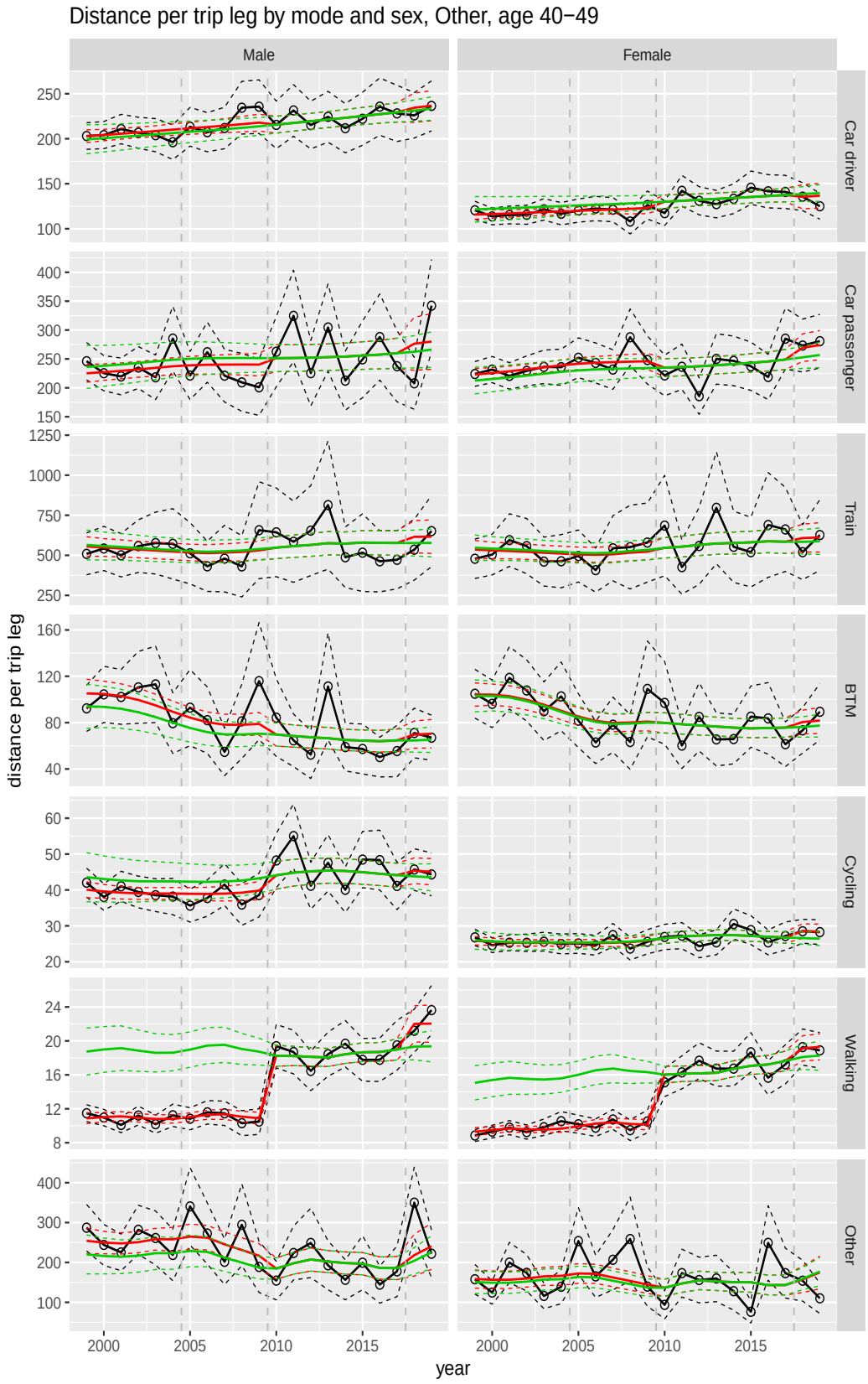

Figure A.134 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

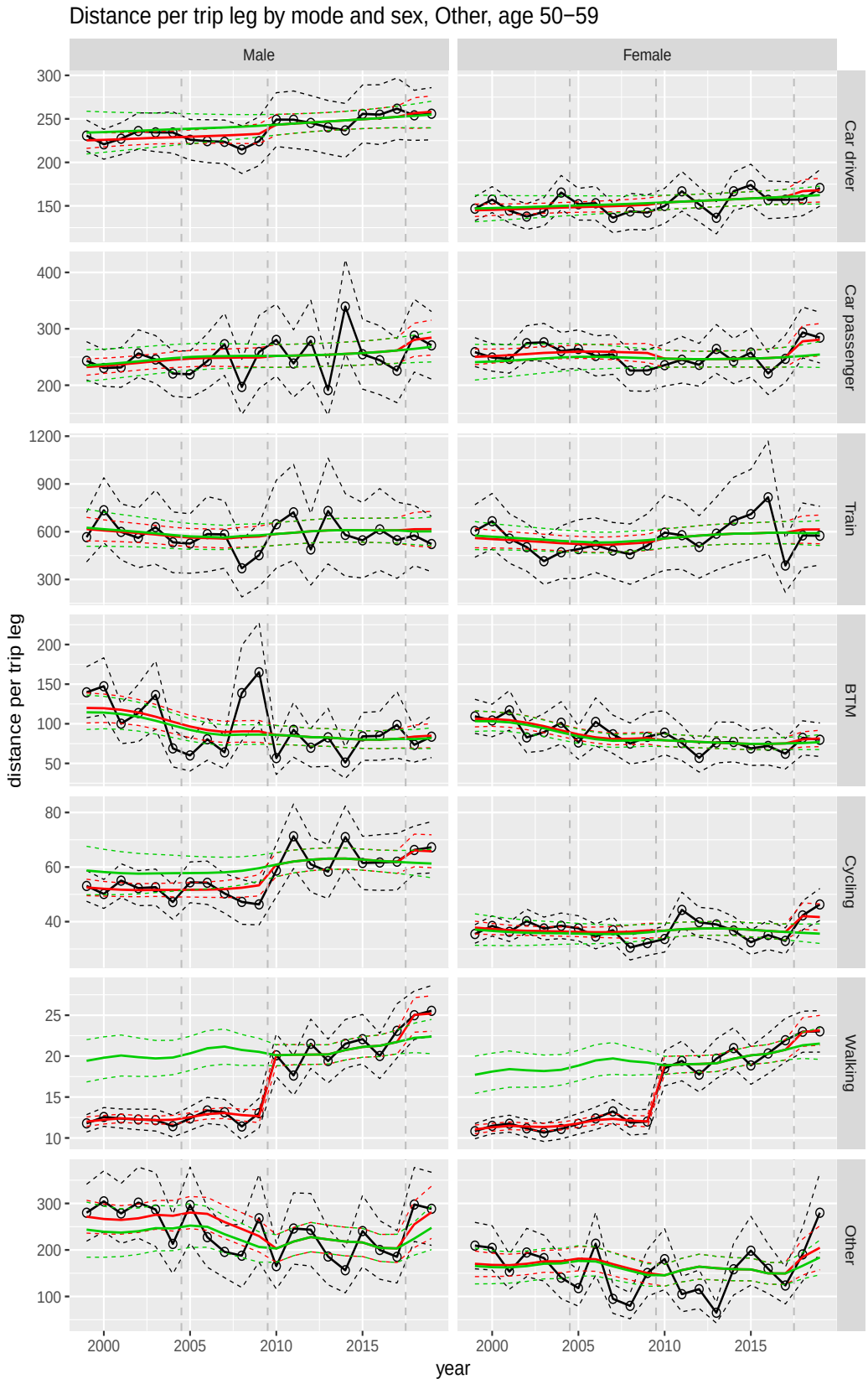

Figure A.135 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

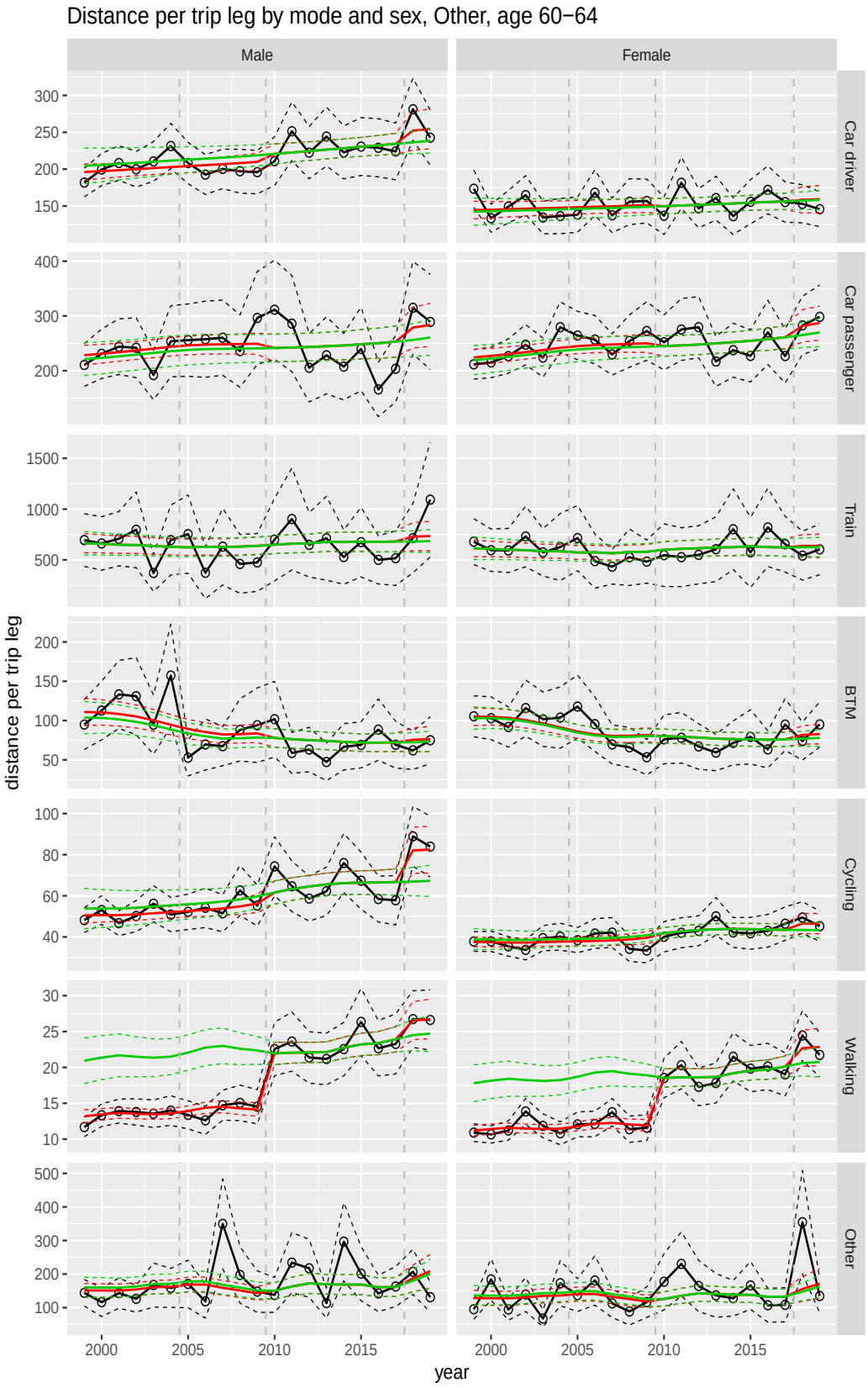

Figure A.136 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

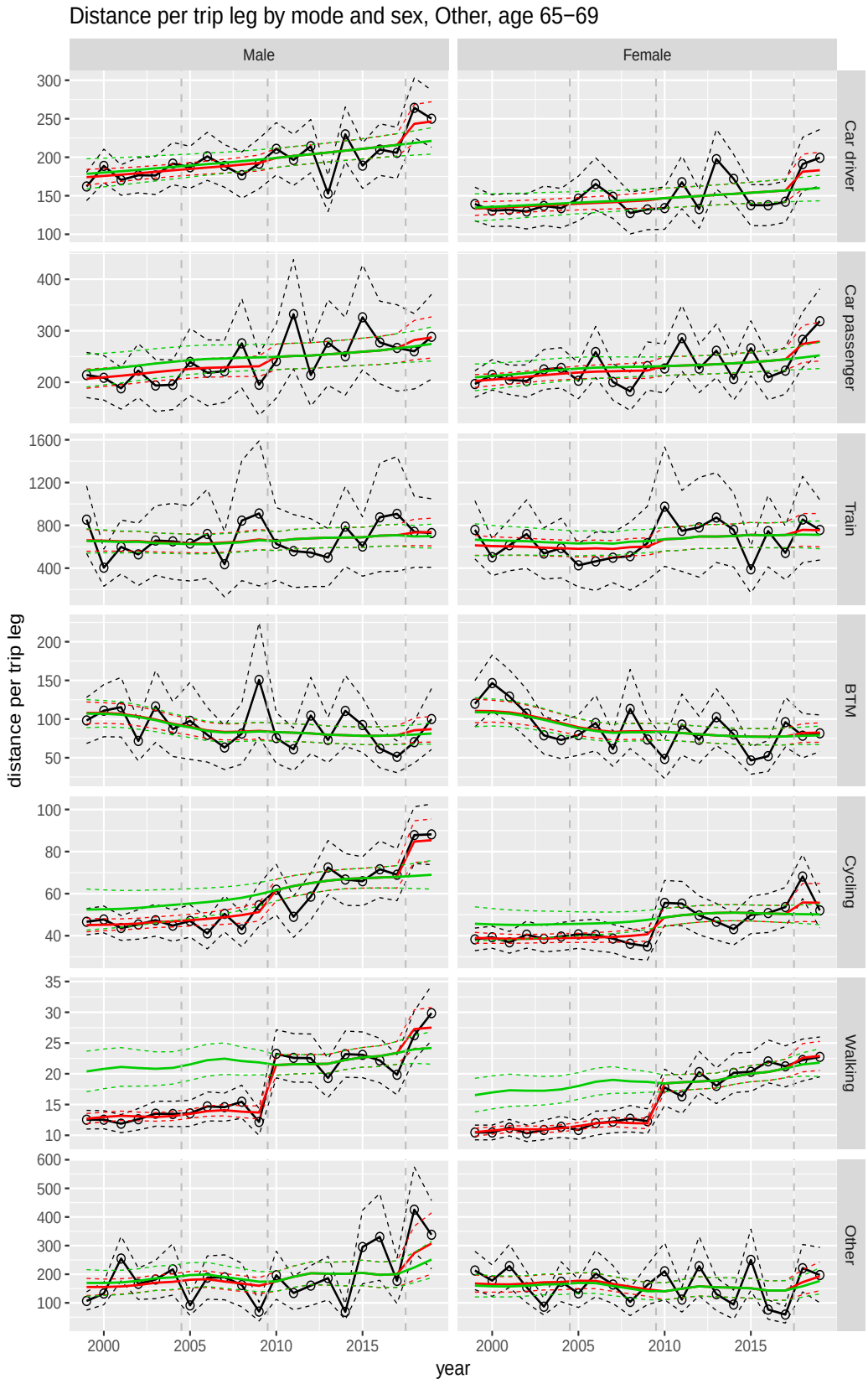

Figure A.137 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

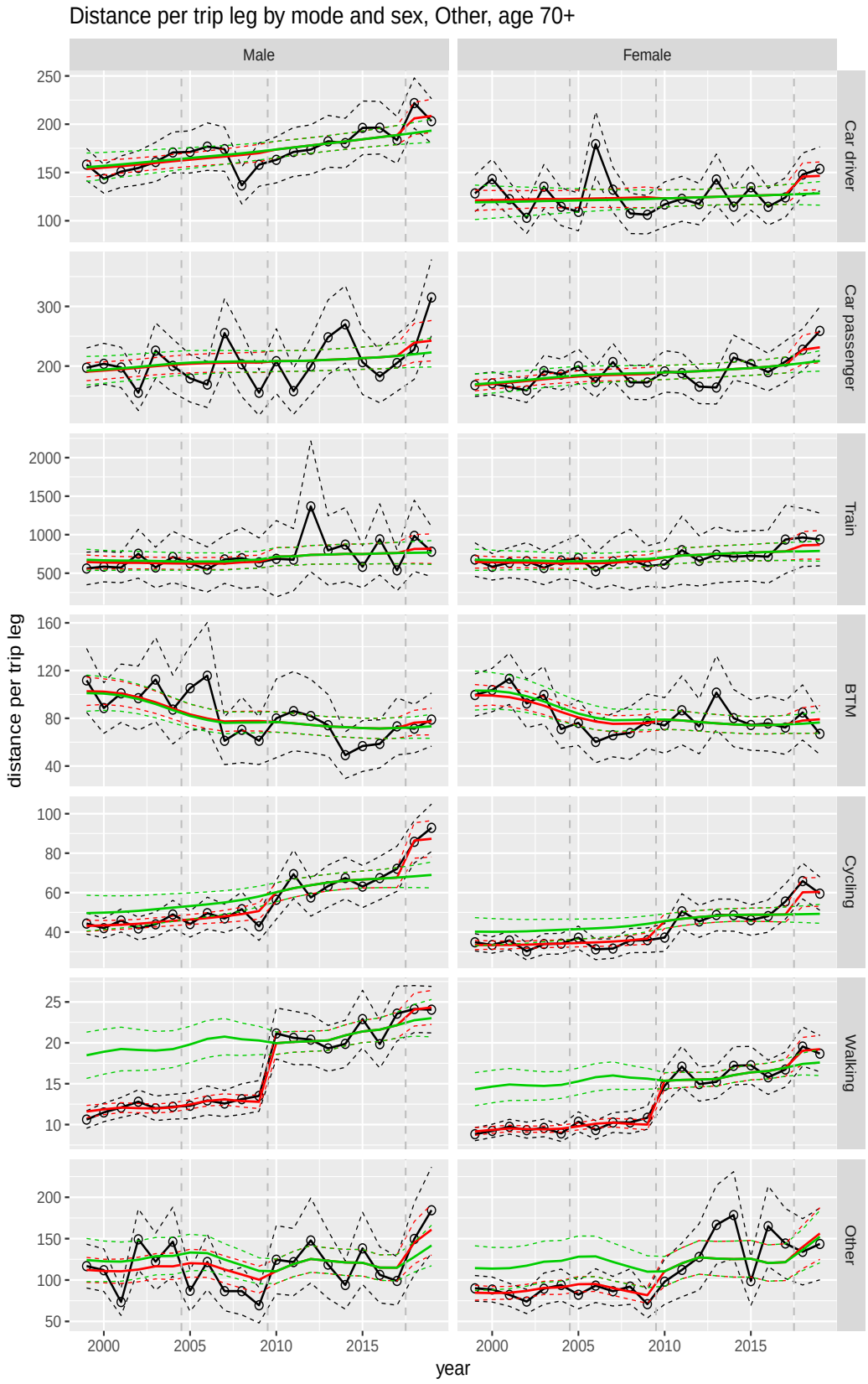

Figure A.138 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

A.3 Average distance per person per day

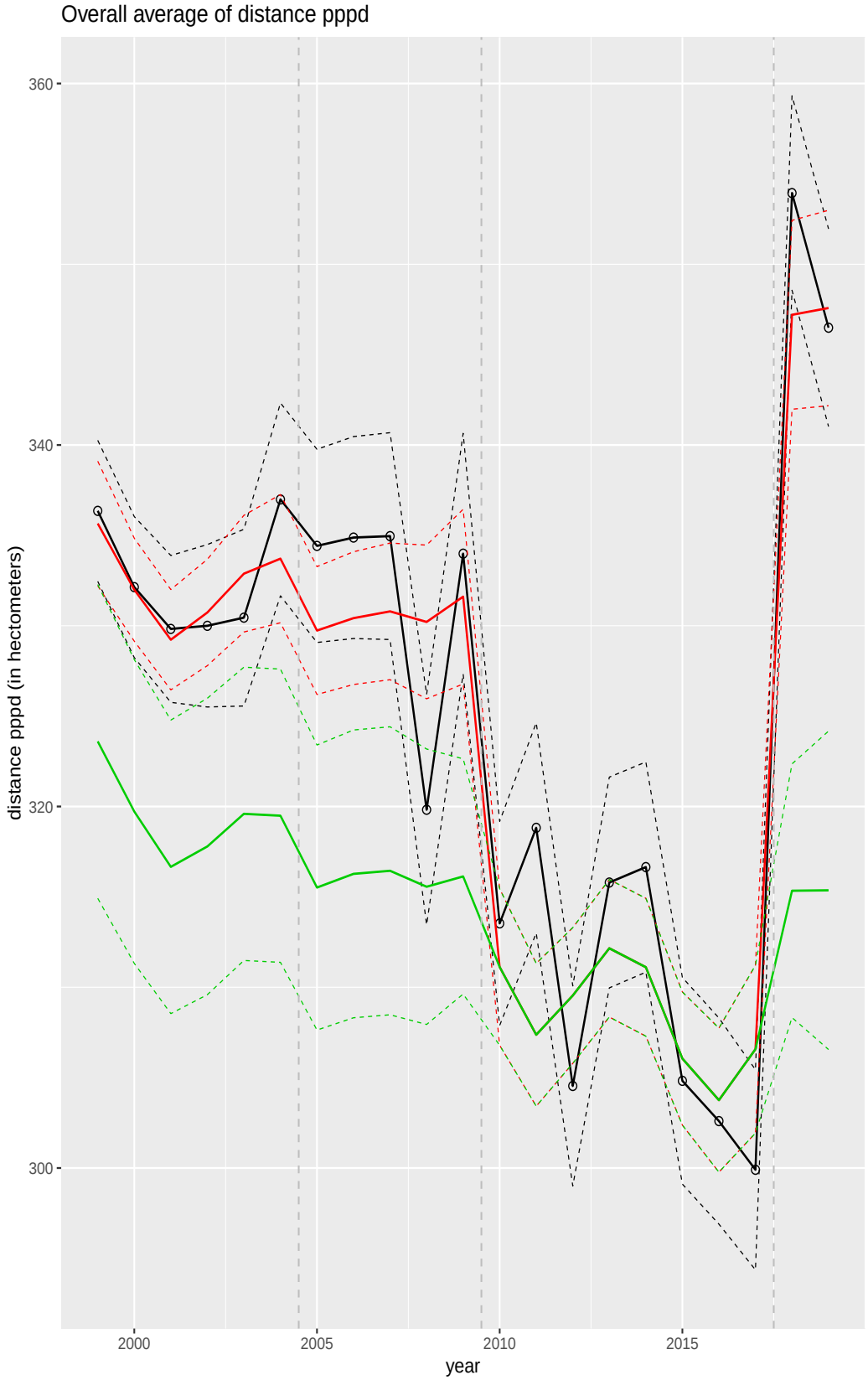

Figure A.139 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

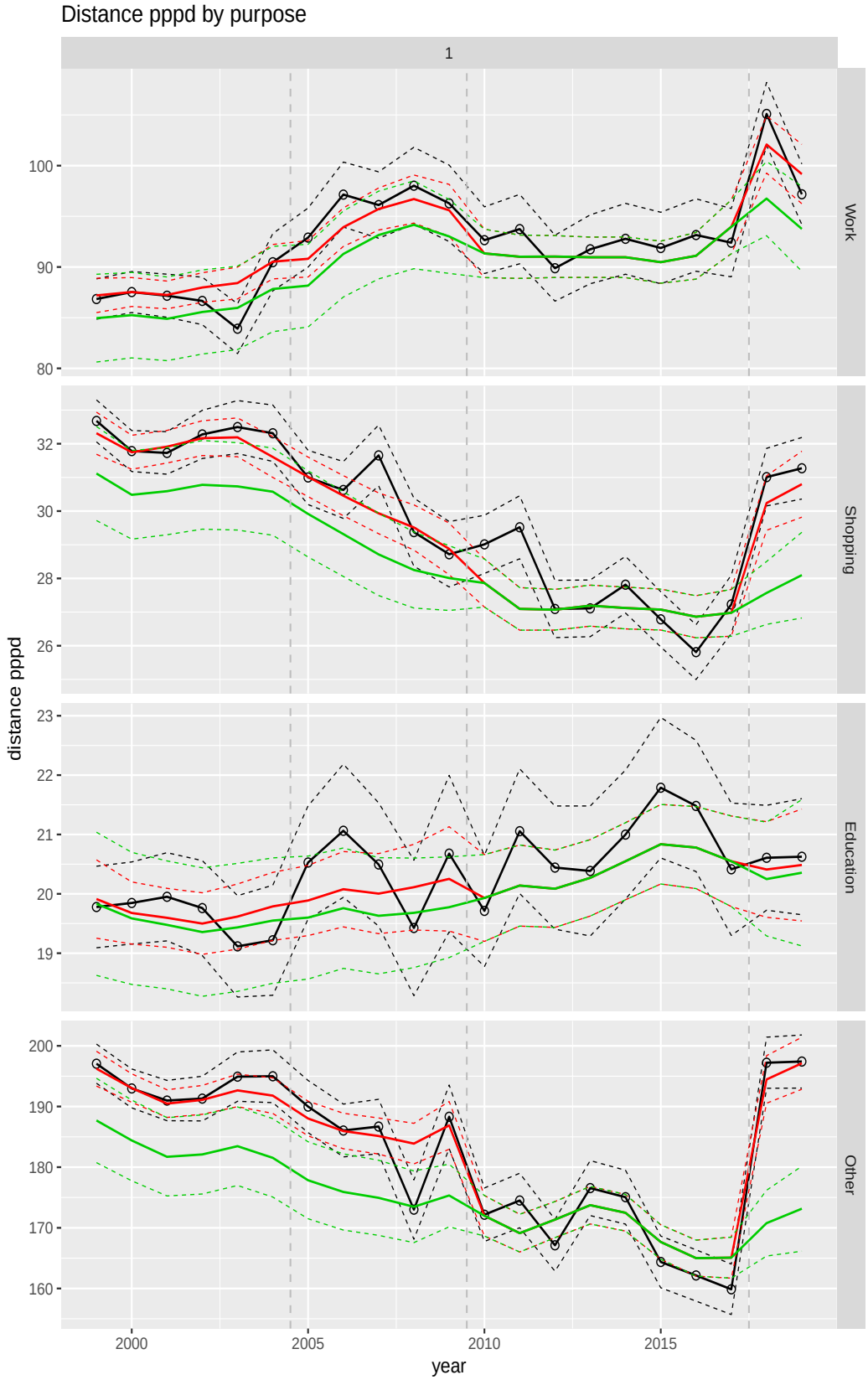

Figure A.140 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

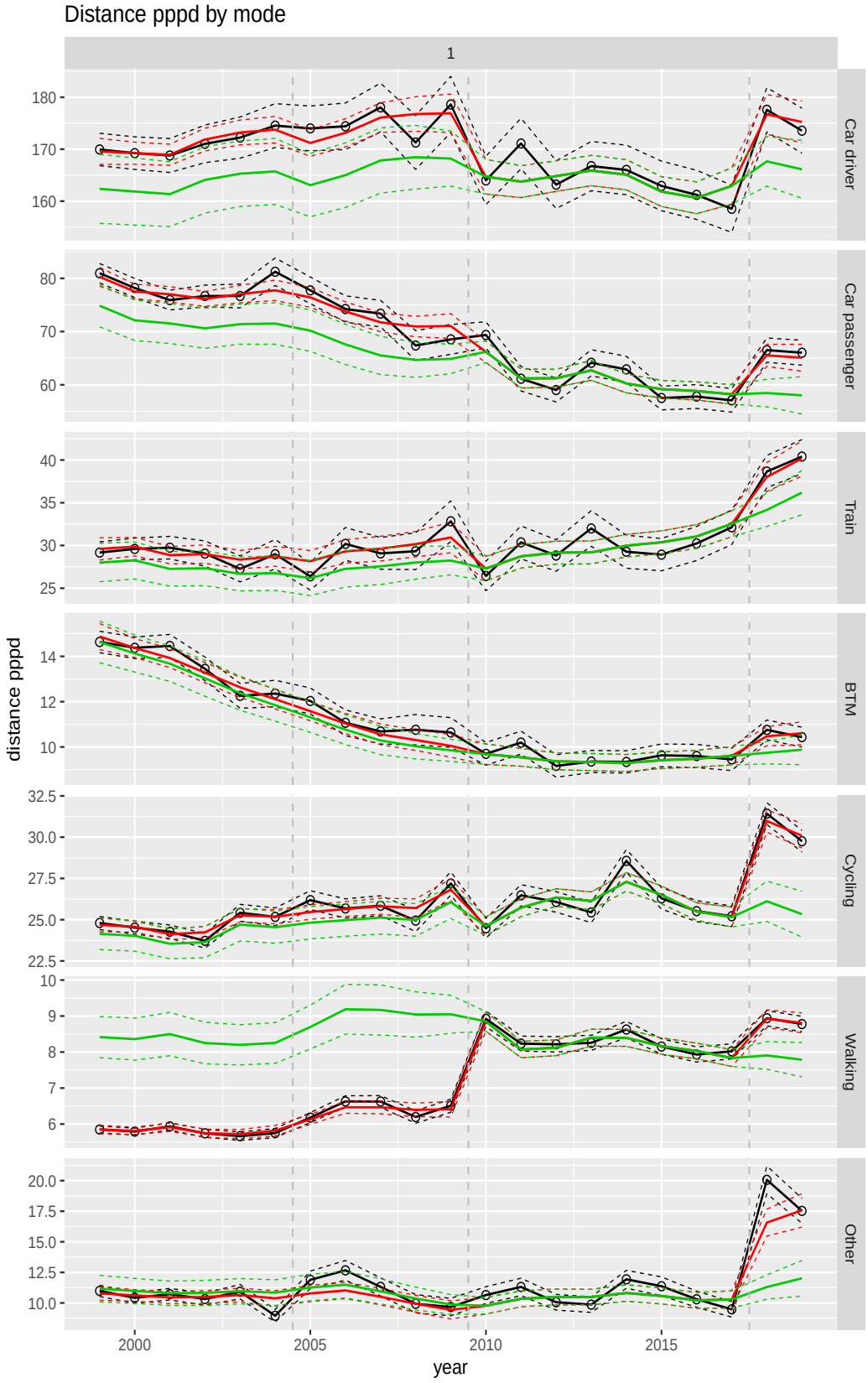

Figure A.141 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

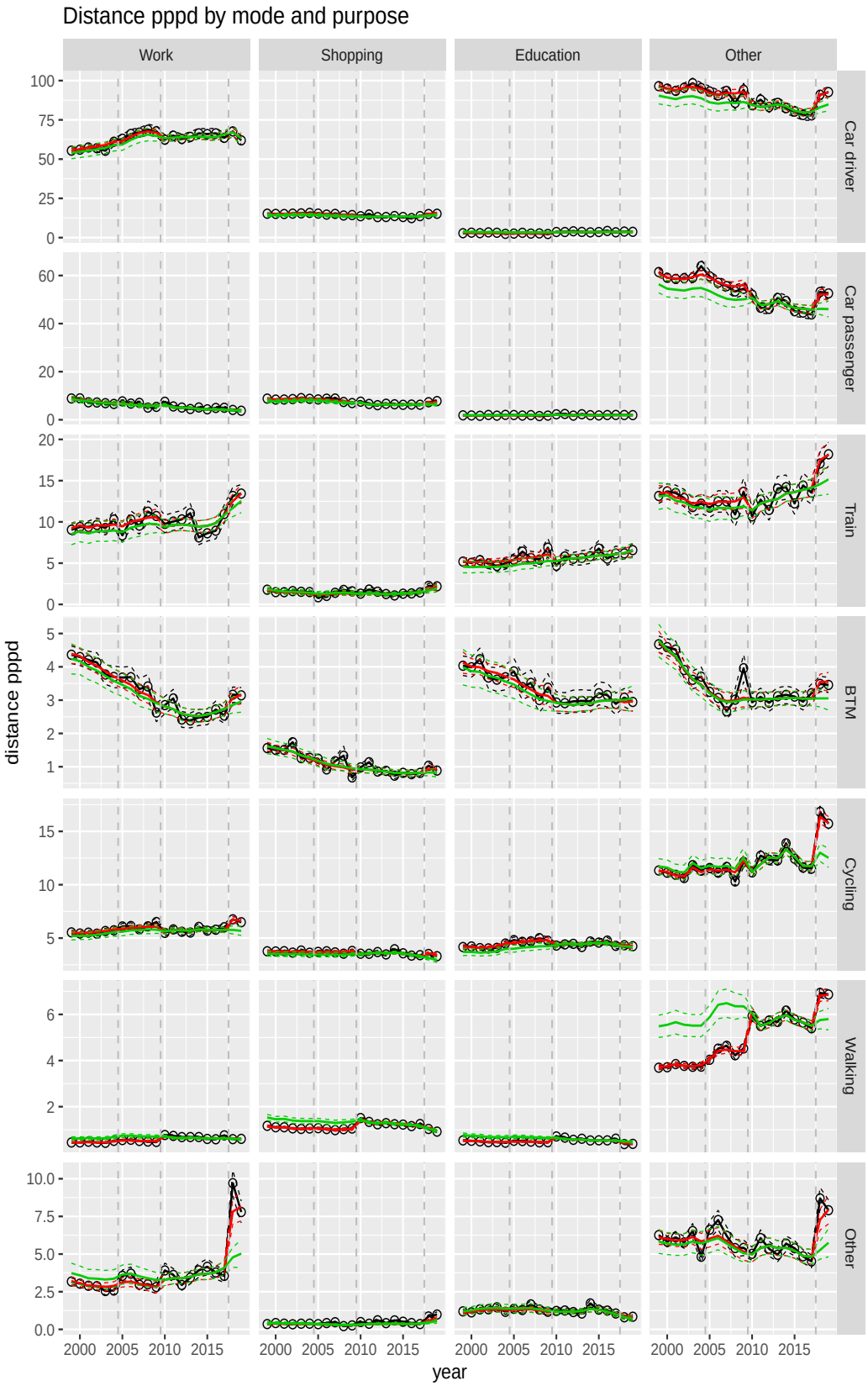

Figure A.142 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

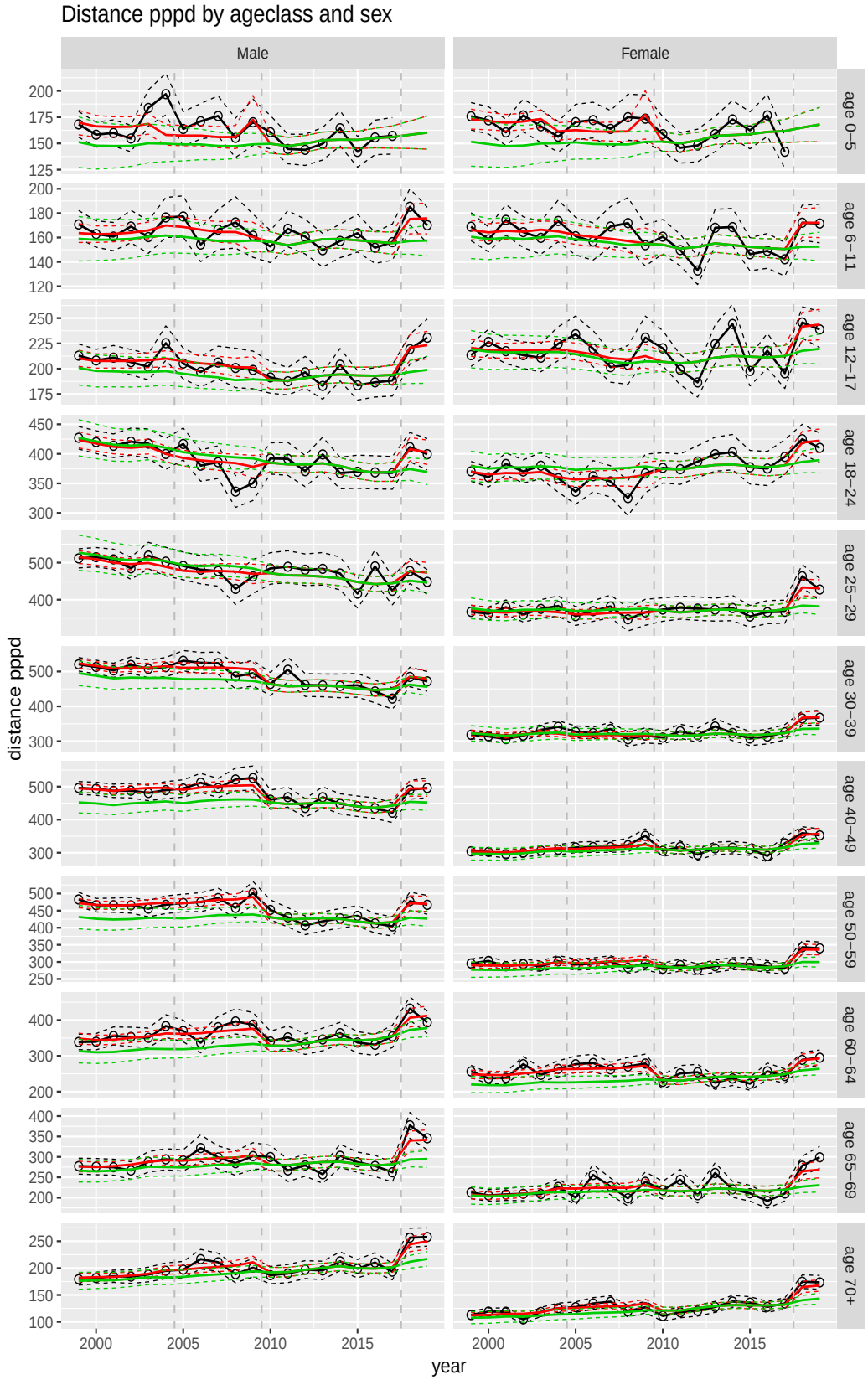

Figure A.143 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

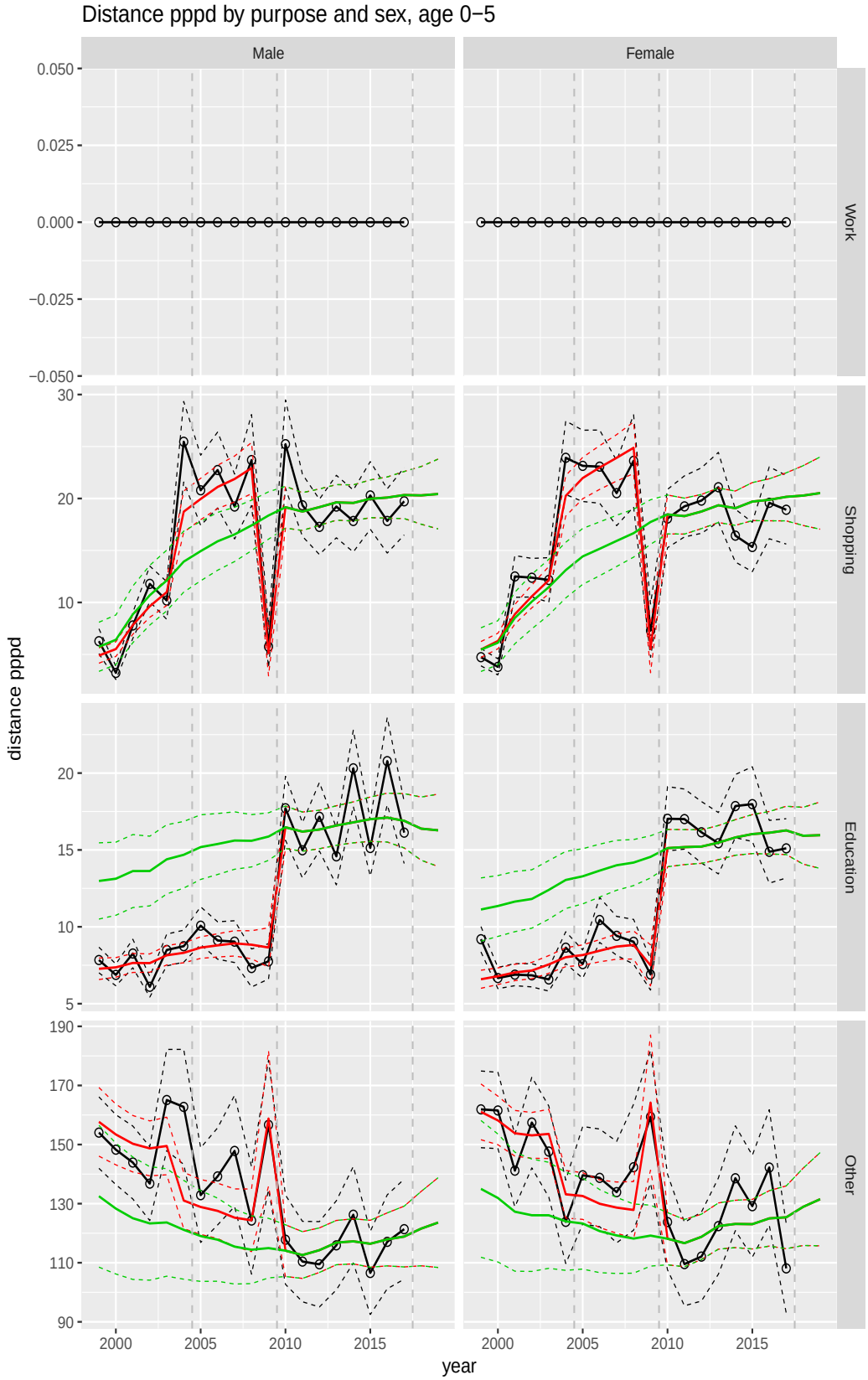

Figure A.144 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

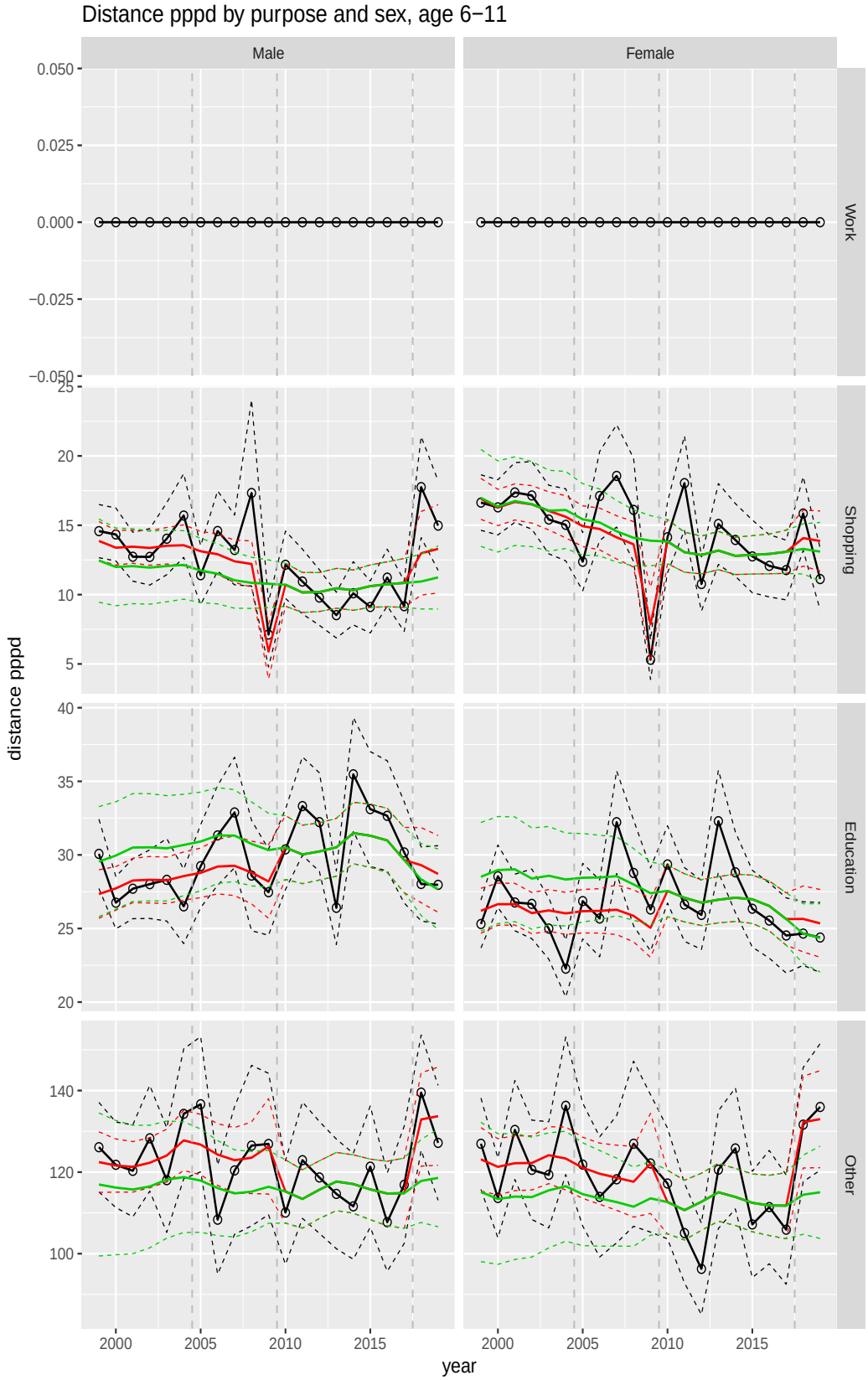

Figure A.145 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

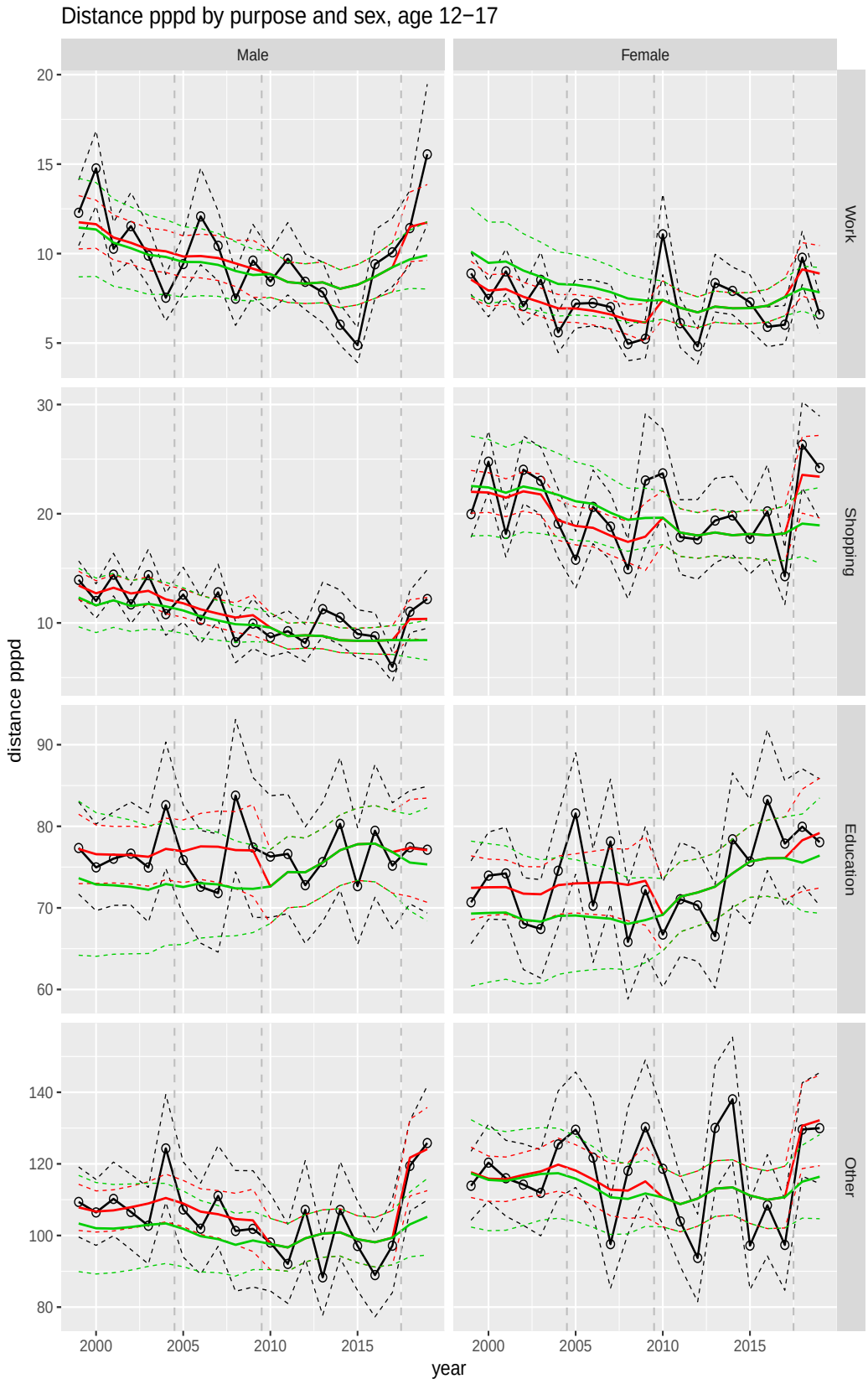

Figure A.146 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

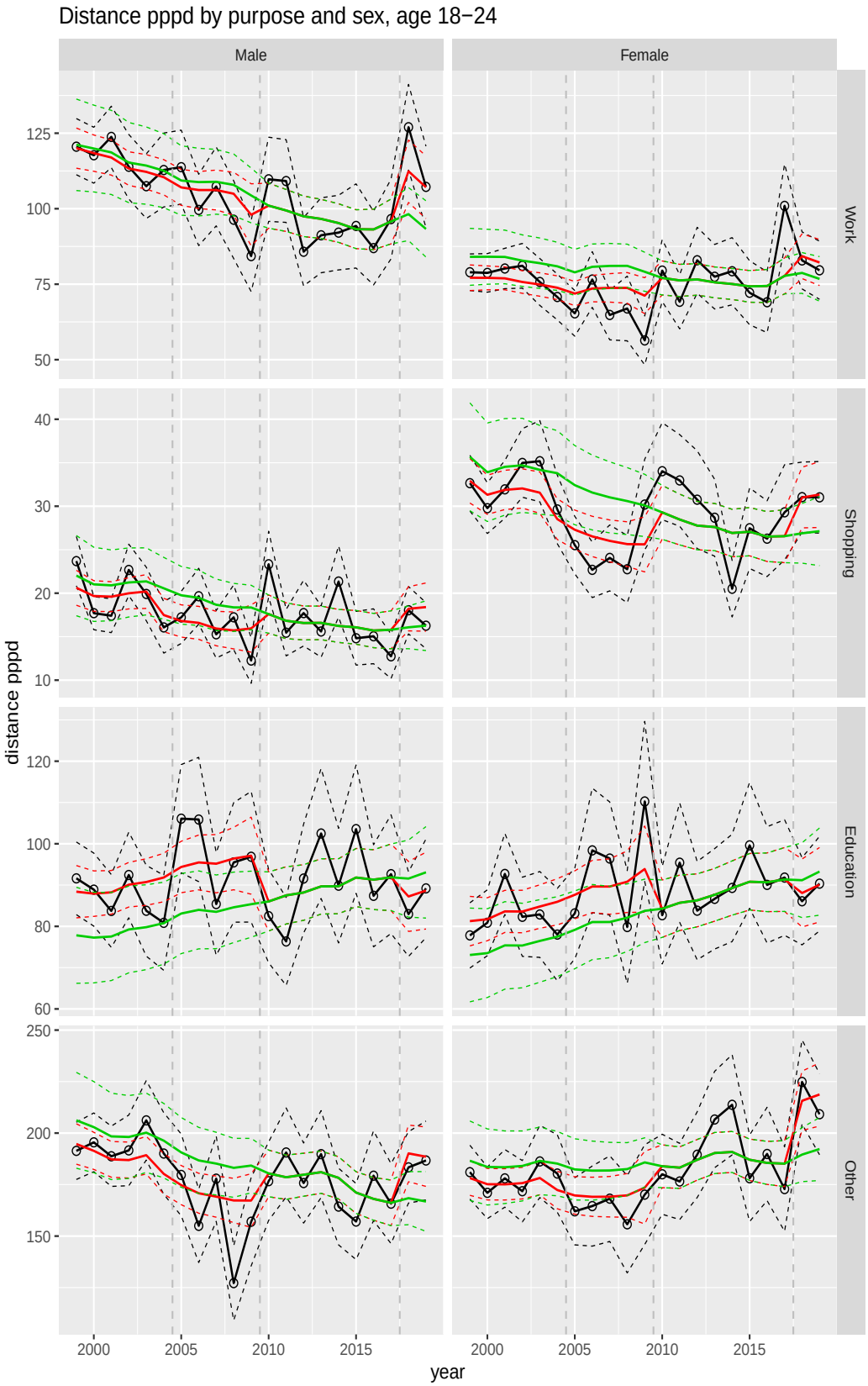

Figure A.147 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

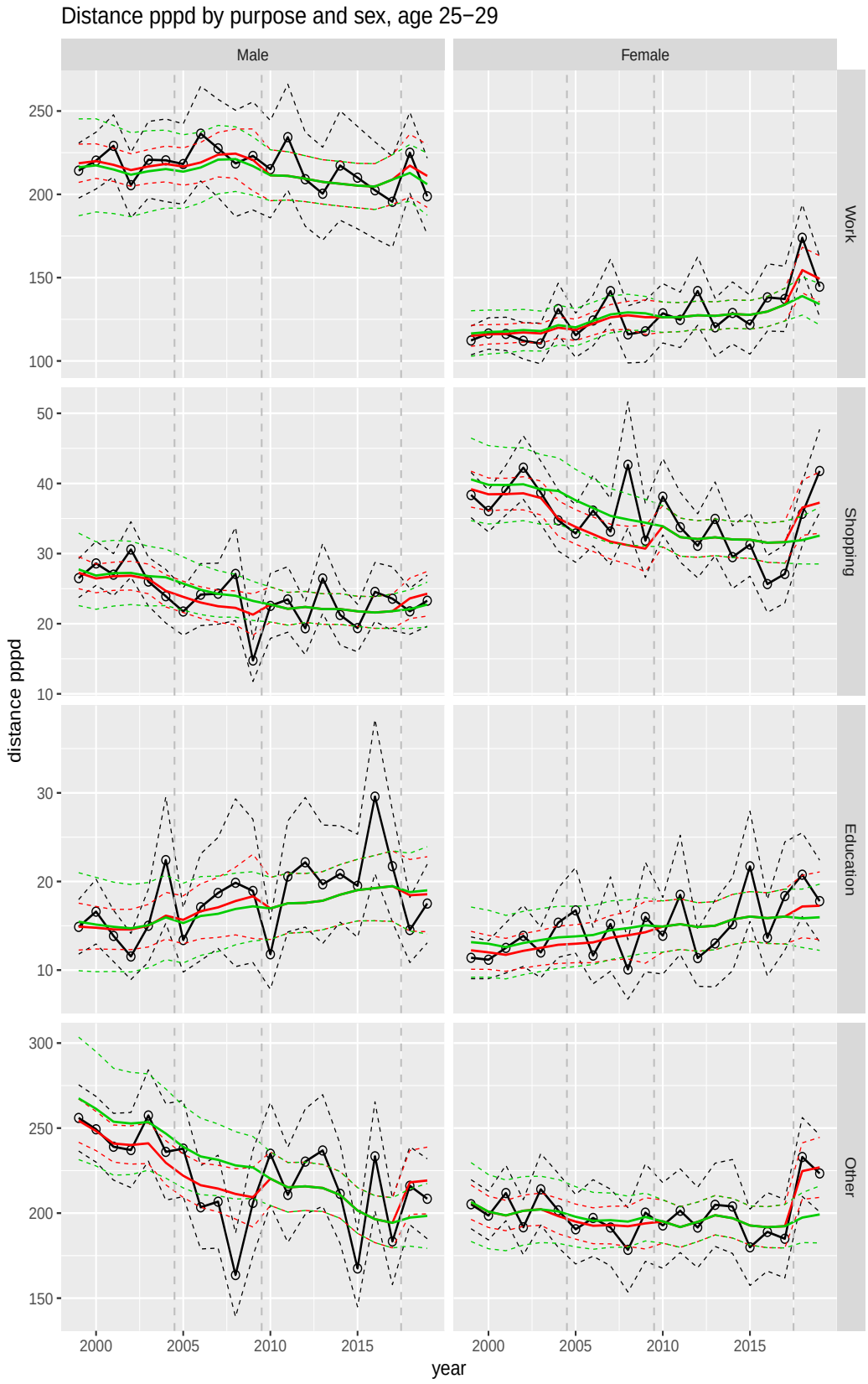

Figure A.148 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

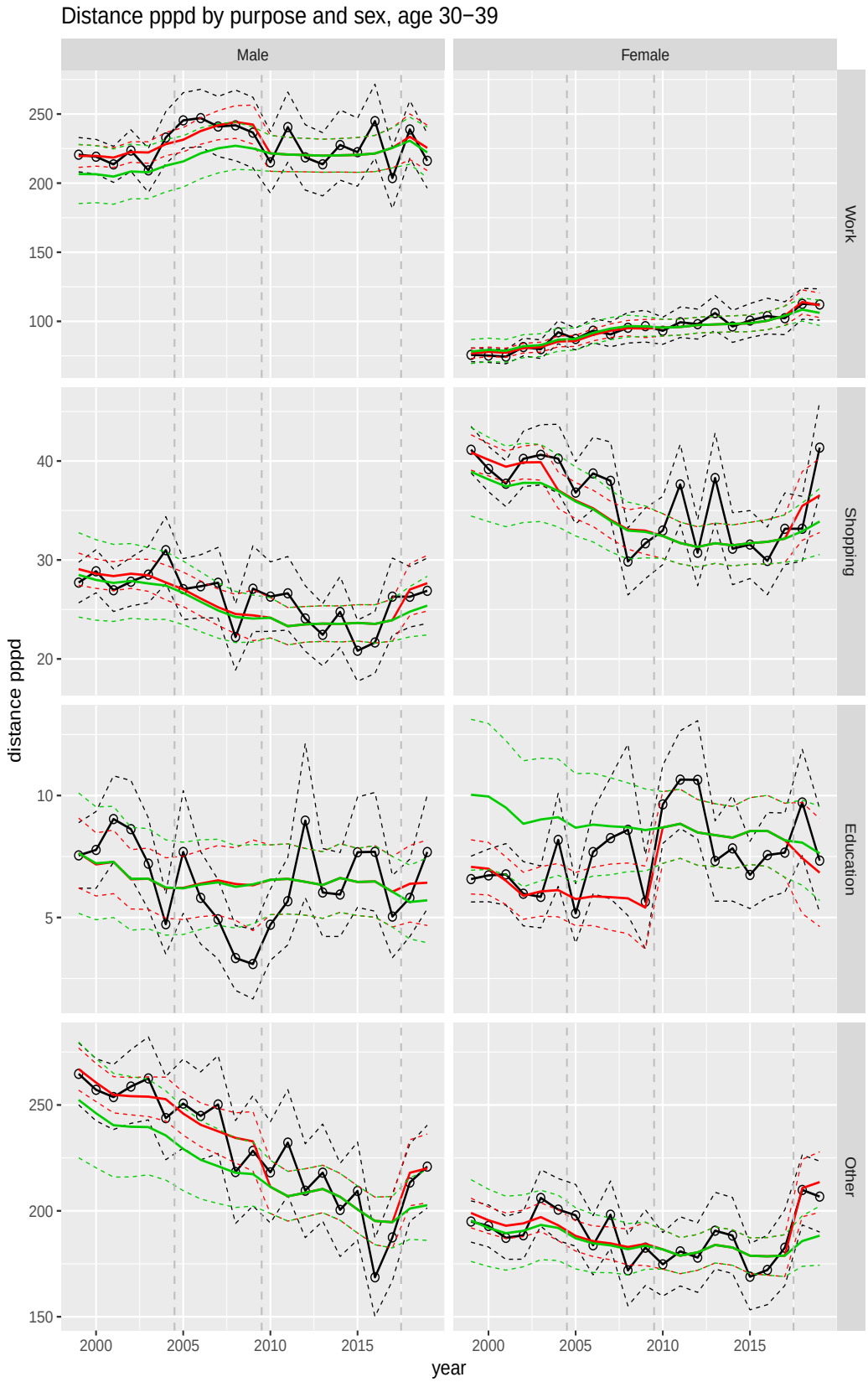

Figure A.149 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

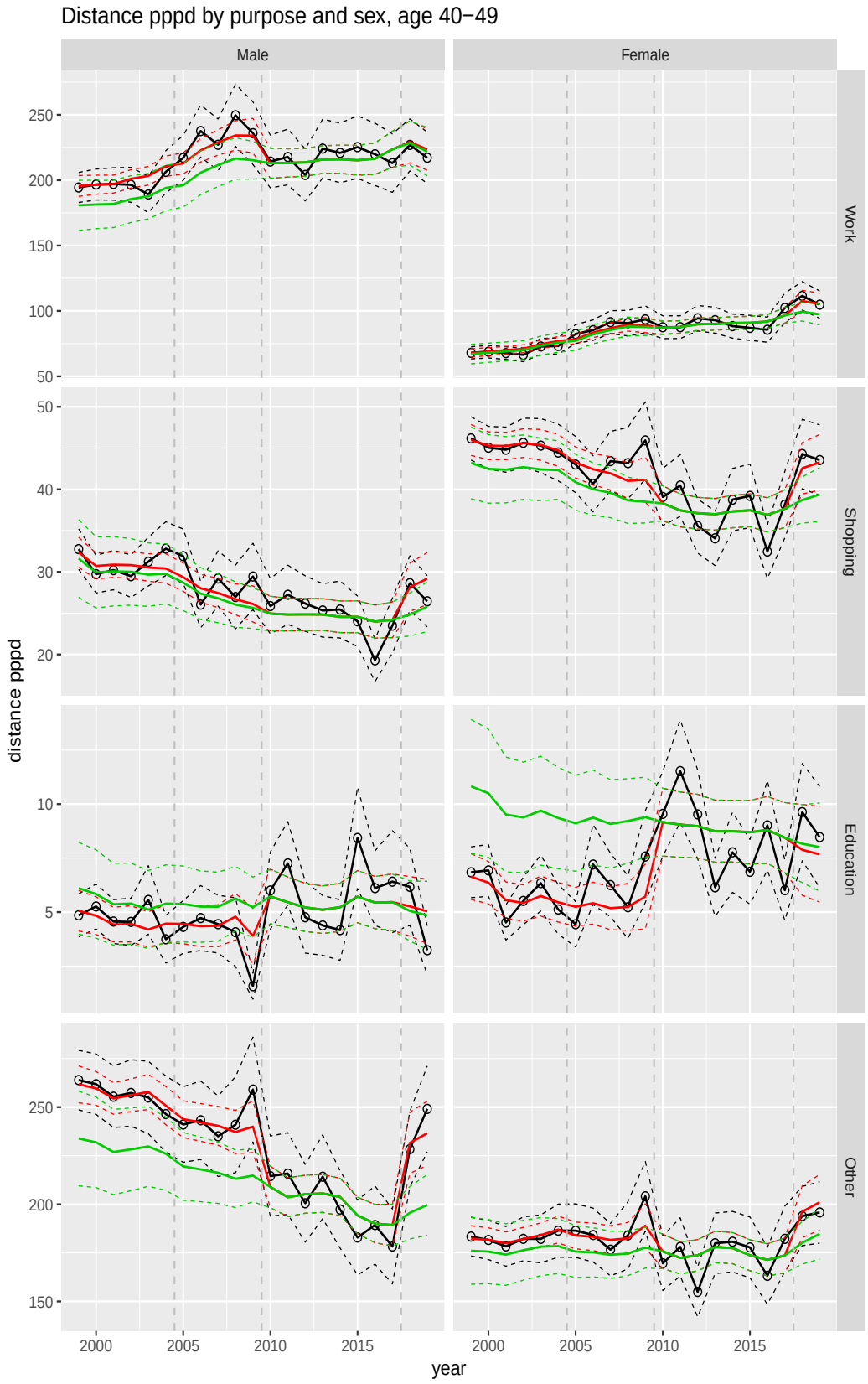

Figure A.150 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

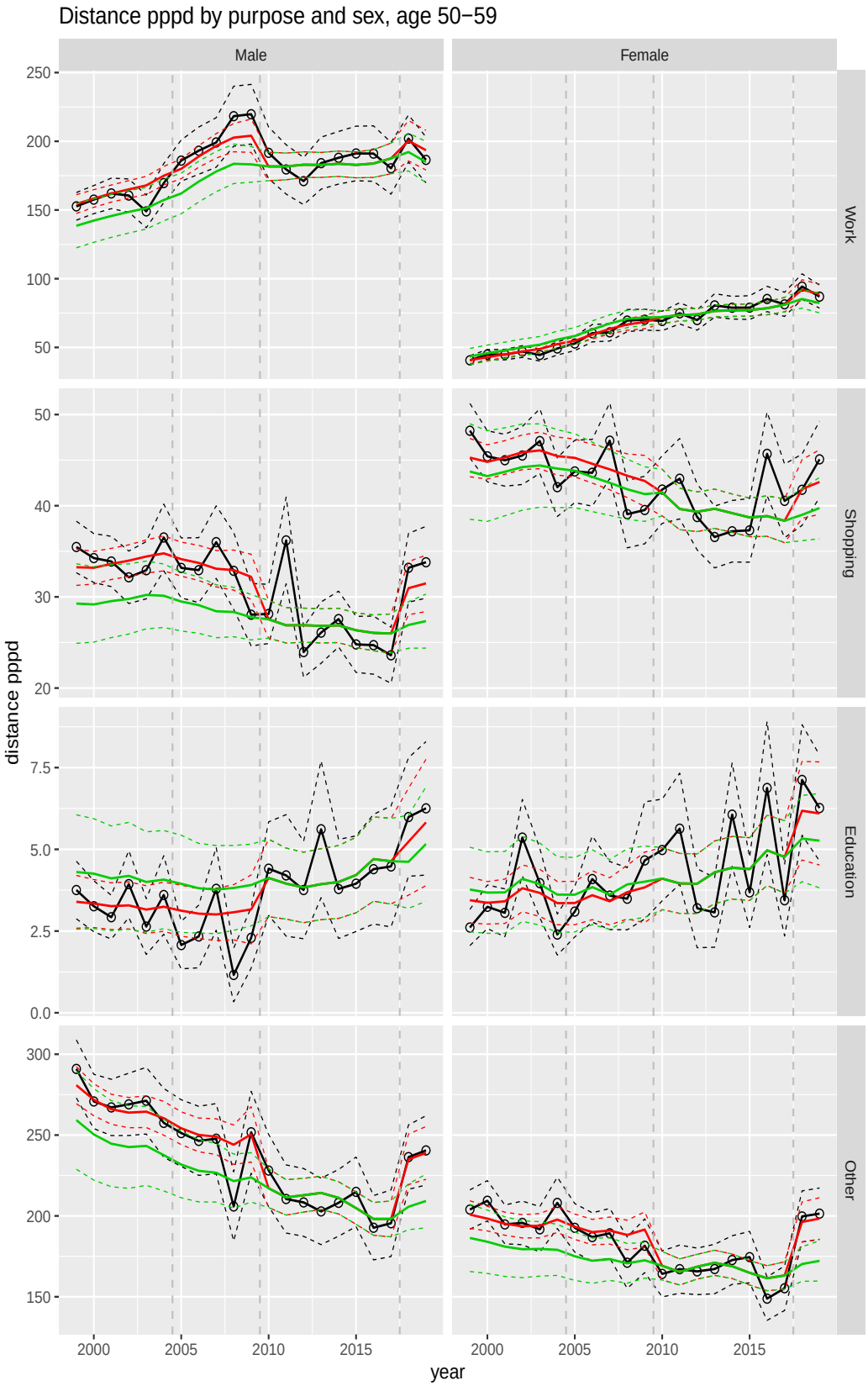

Figure A.151 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

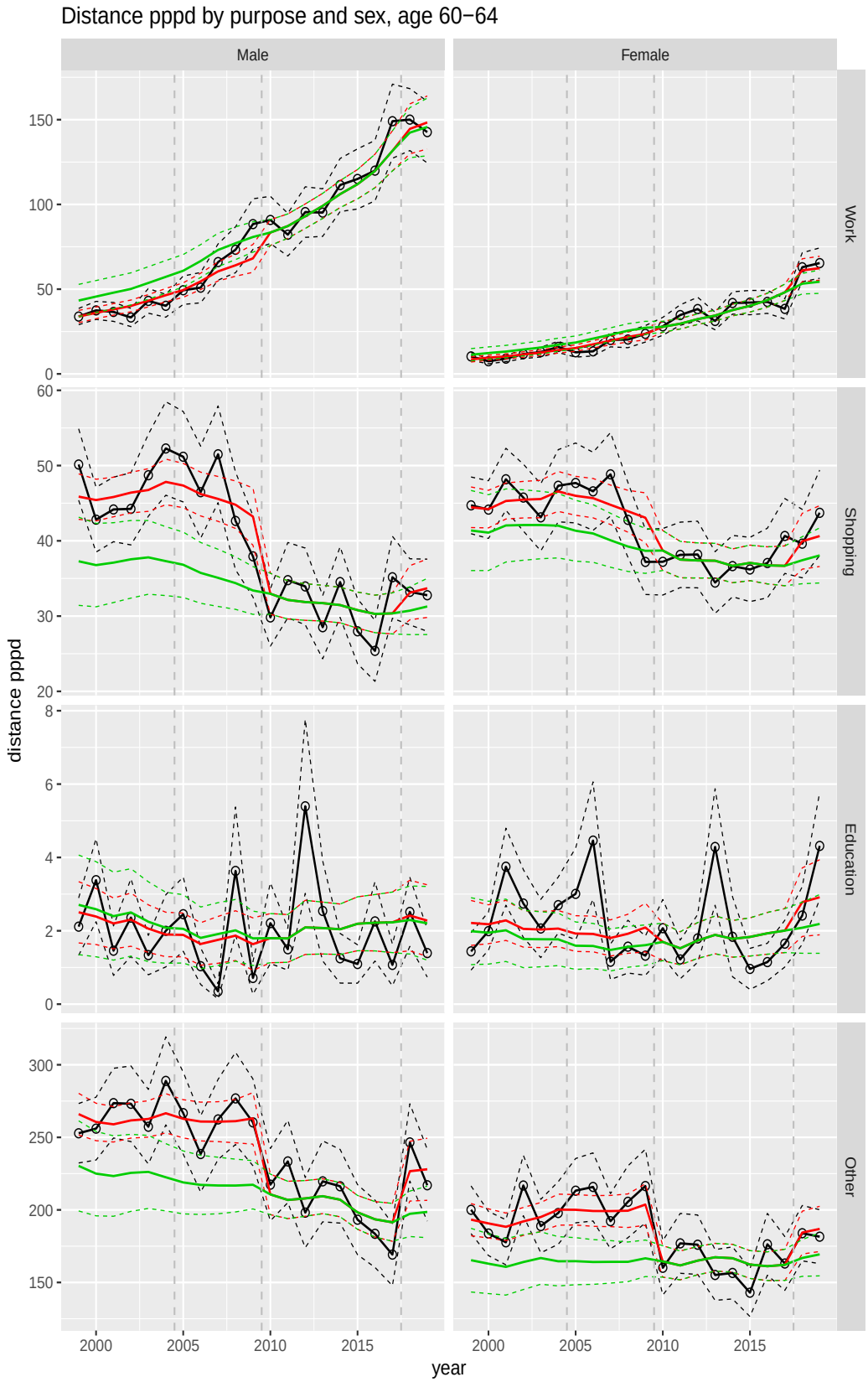

Figure A.152 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

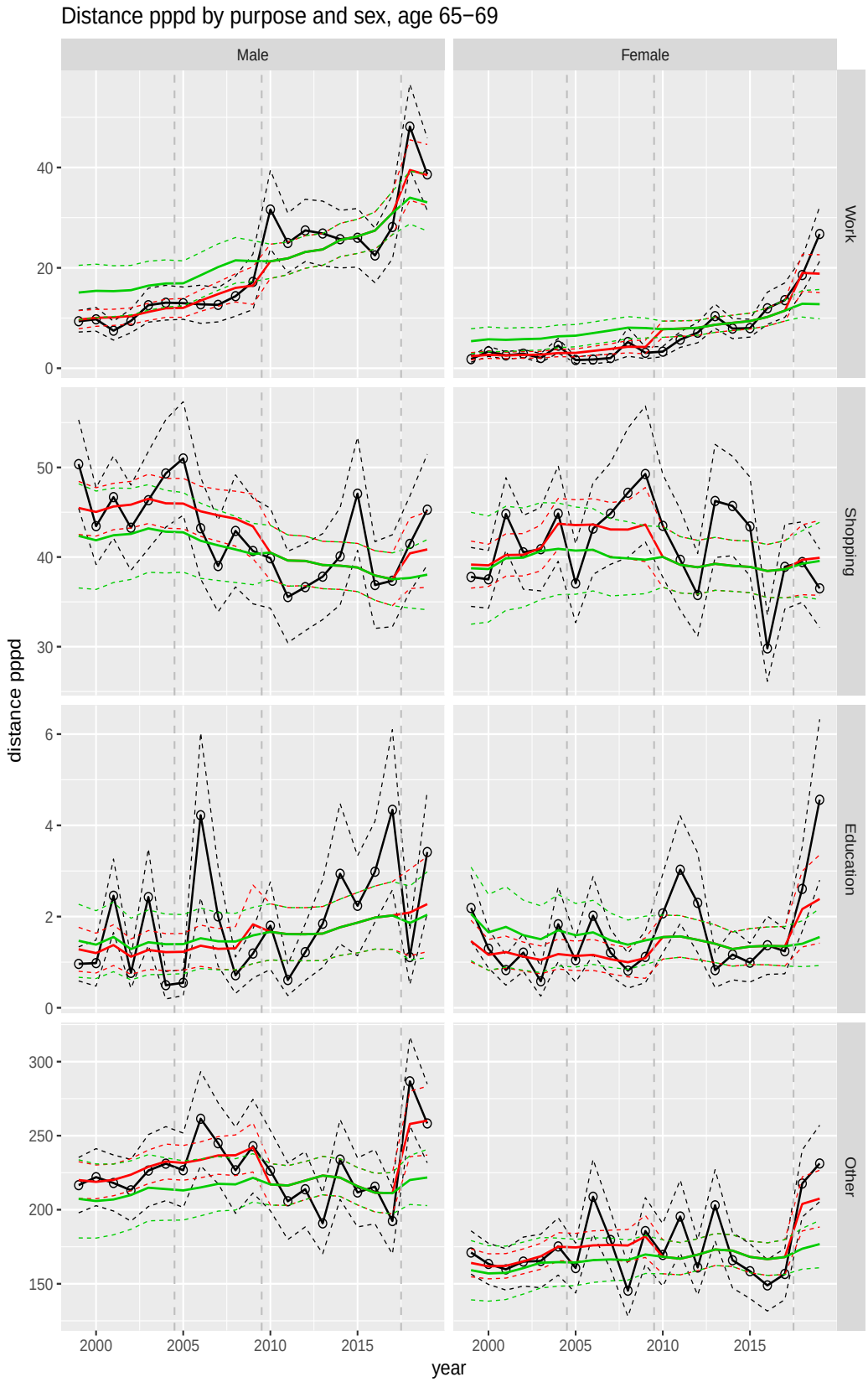

Figure A.153 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

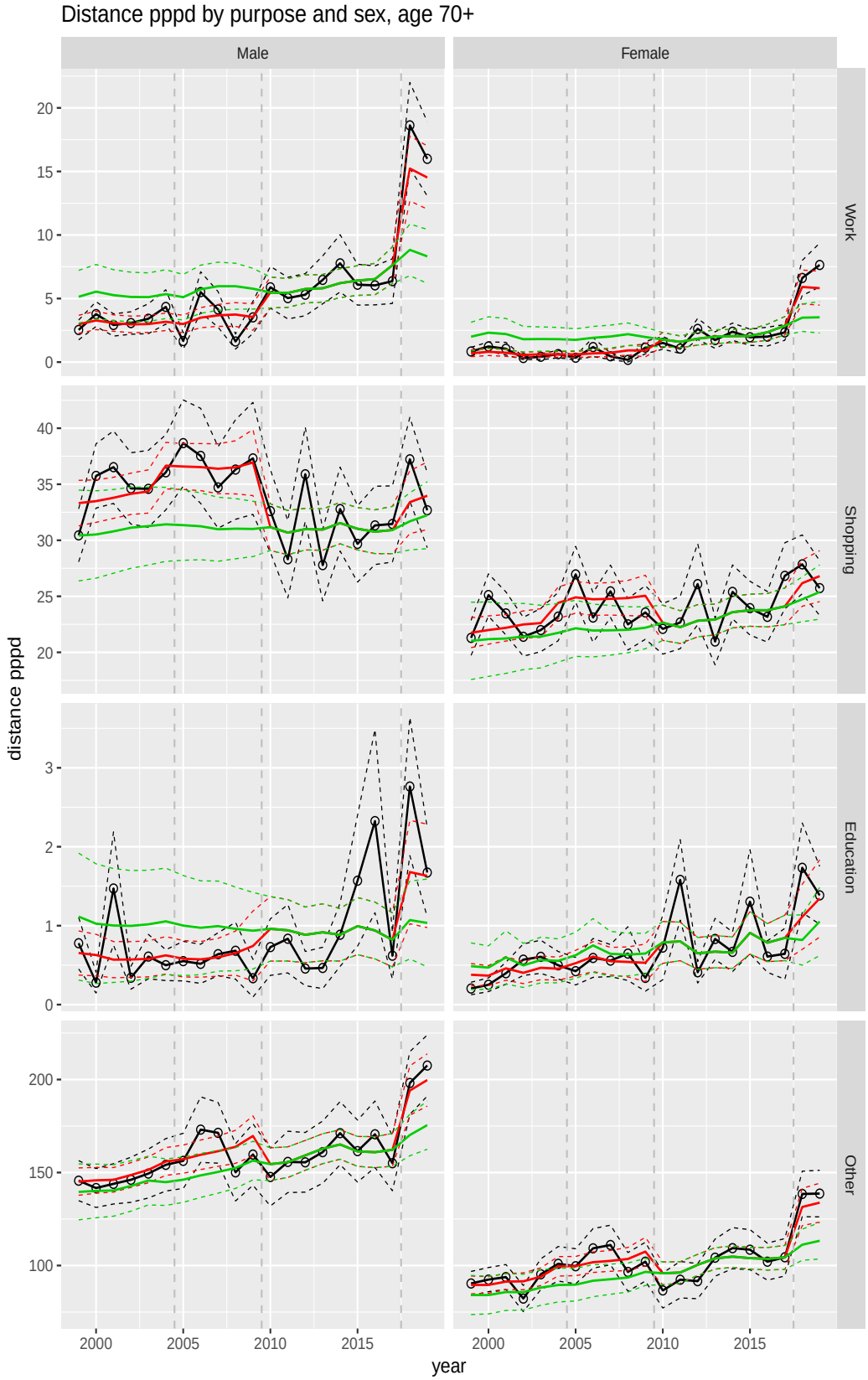

Figure A.154 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

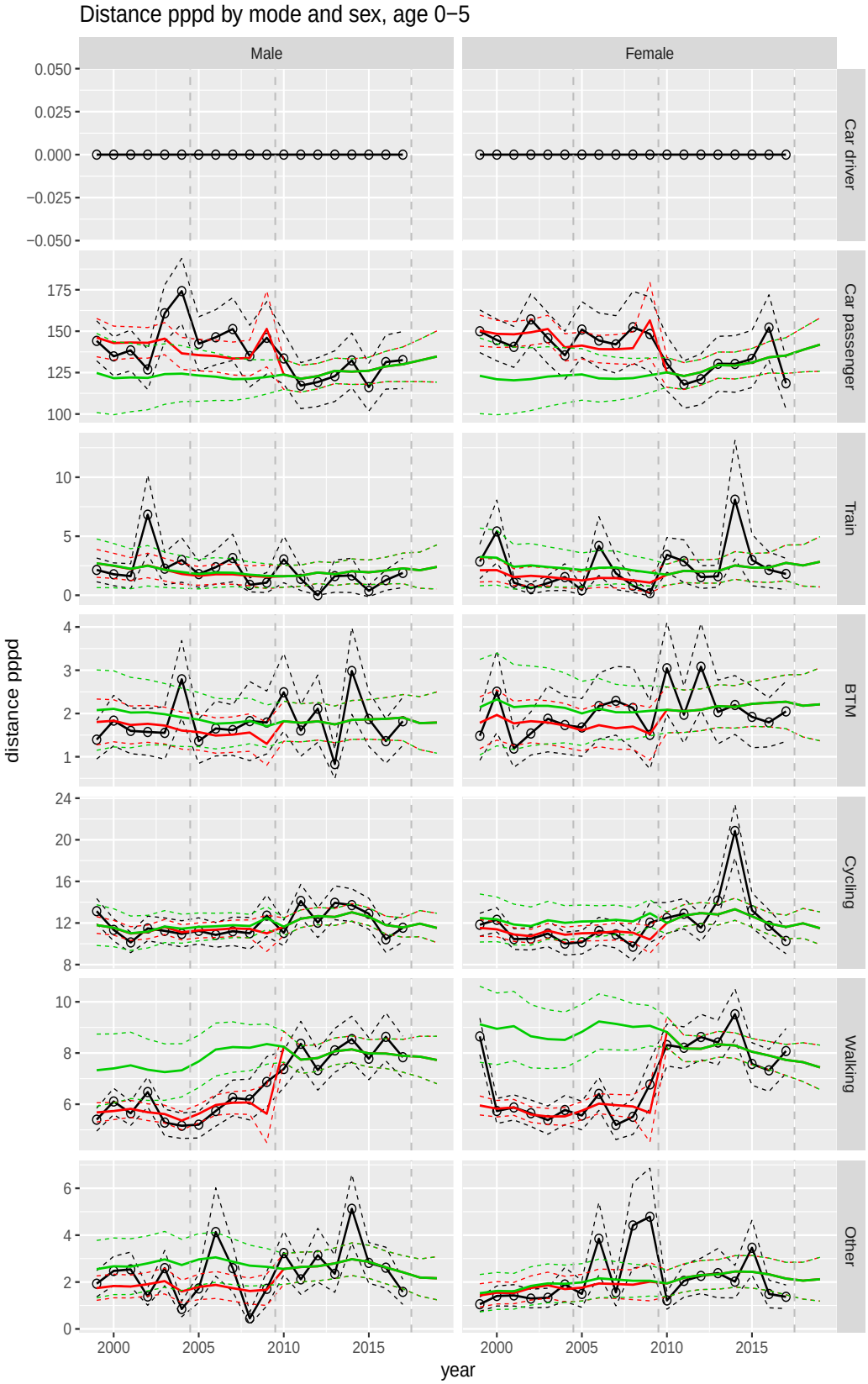

Figure A.155 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

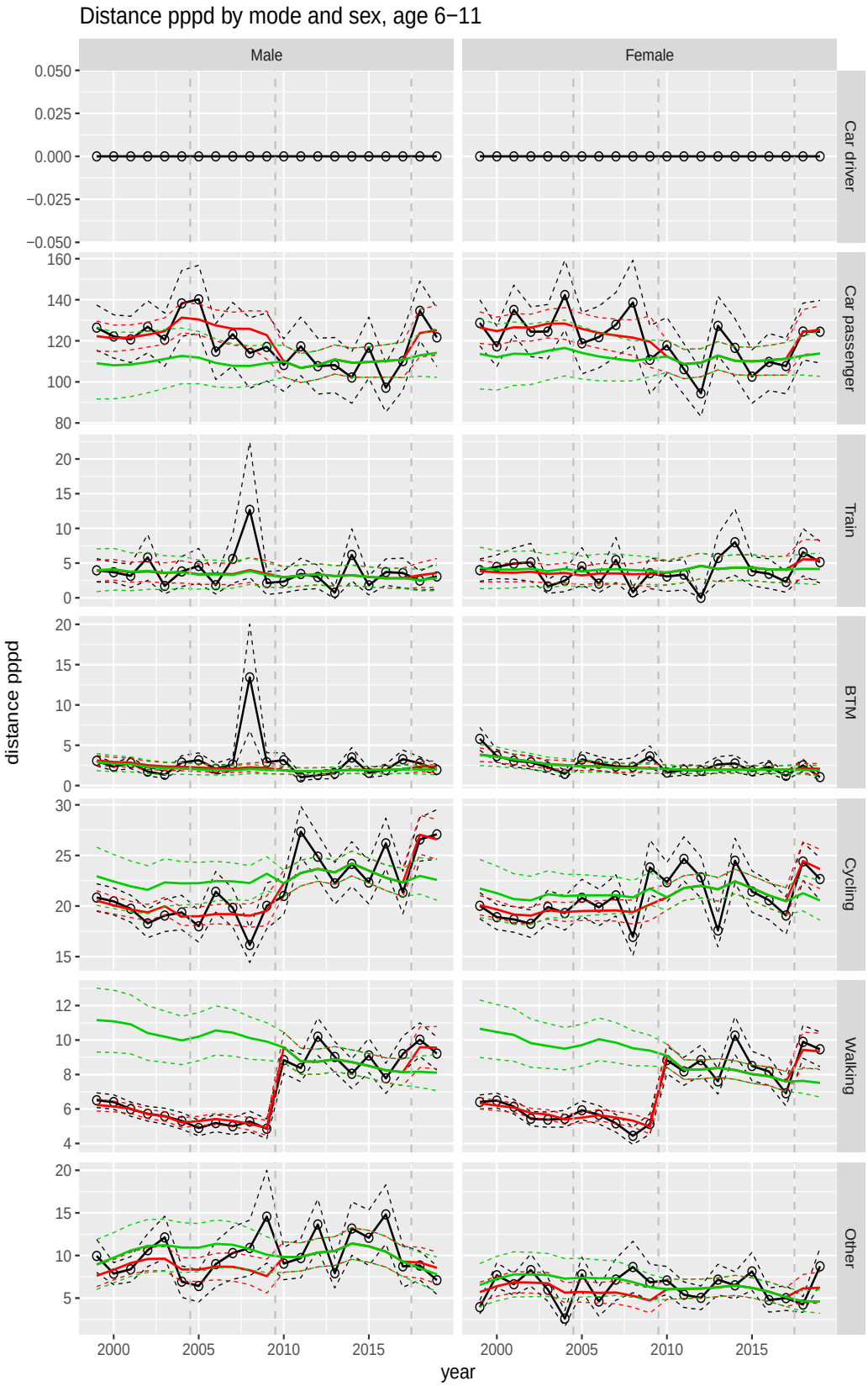

Figure A.156 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

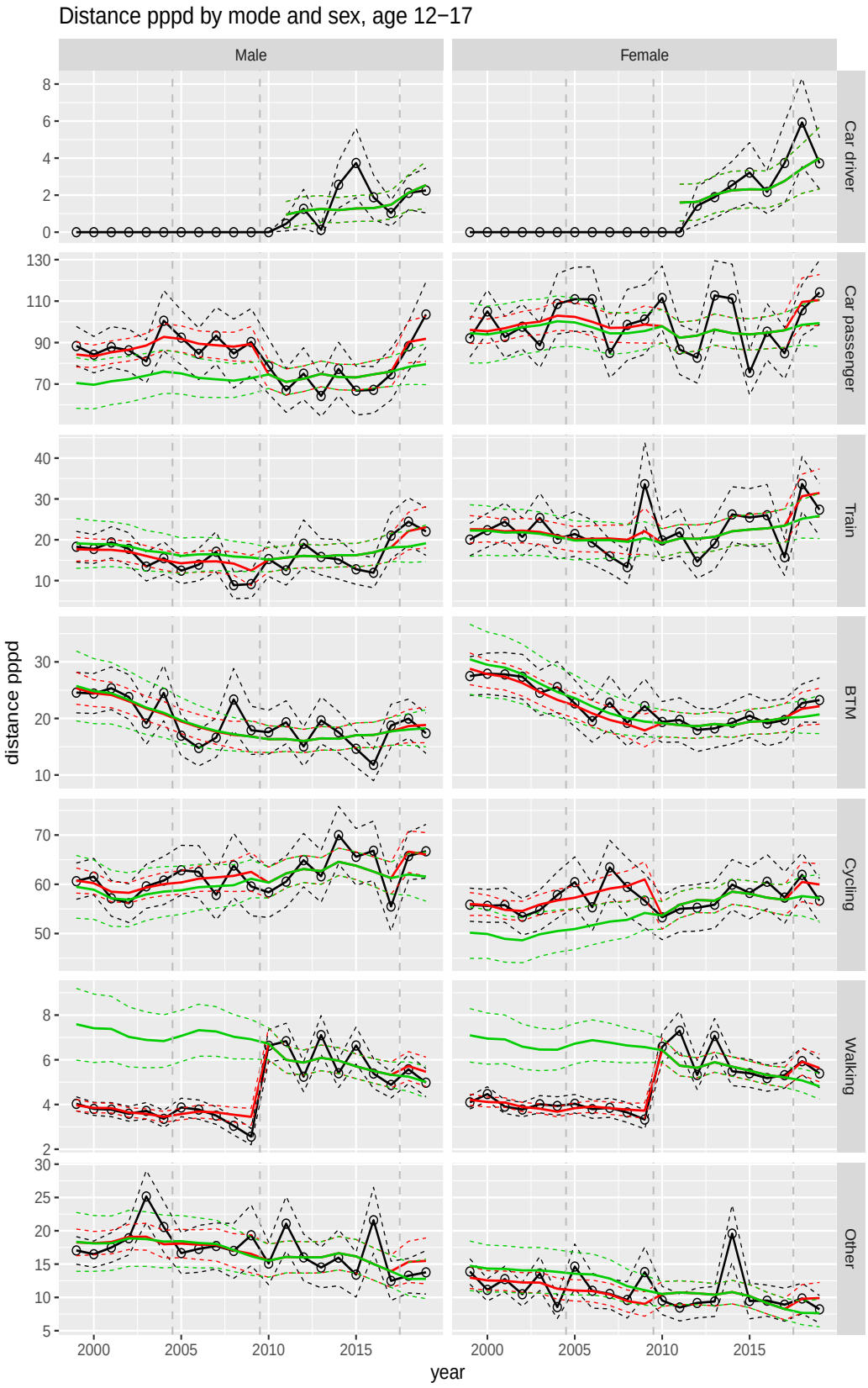

Figure A.157 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

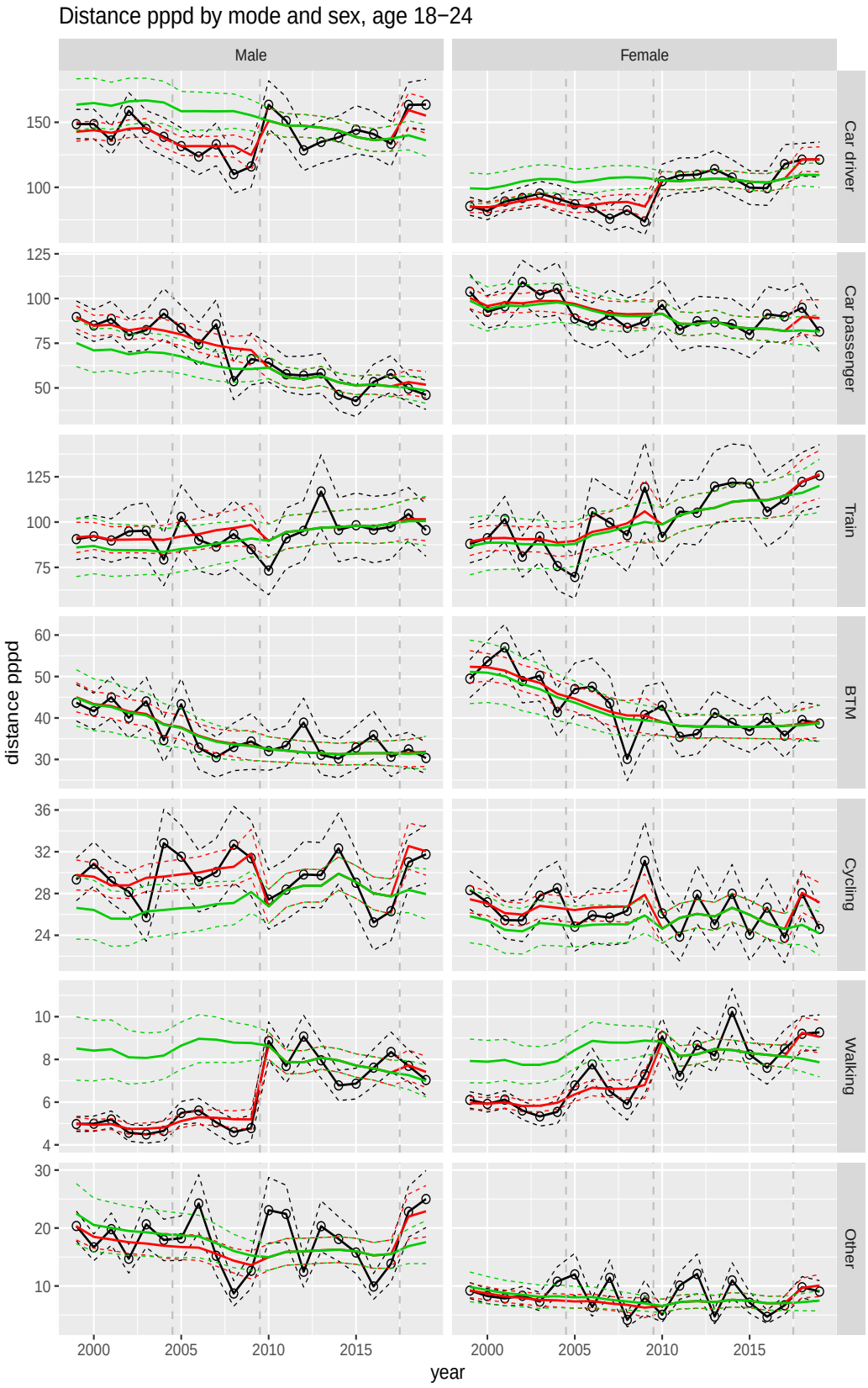

Figure A.158 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

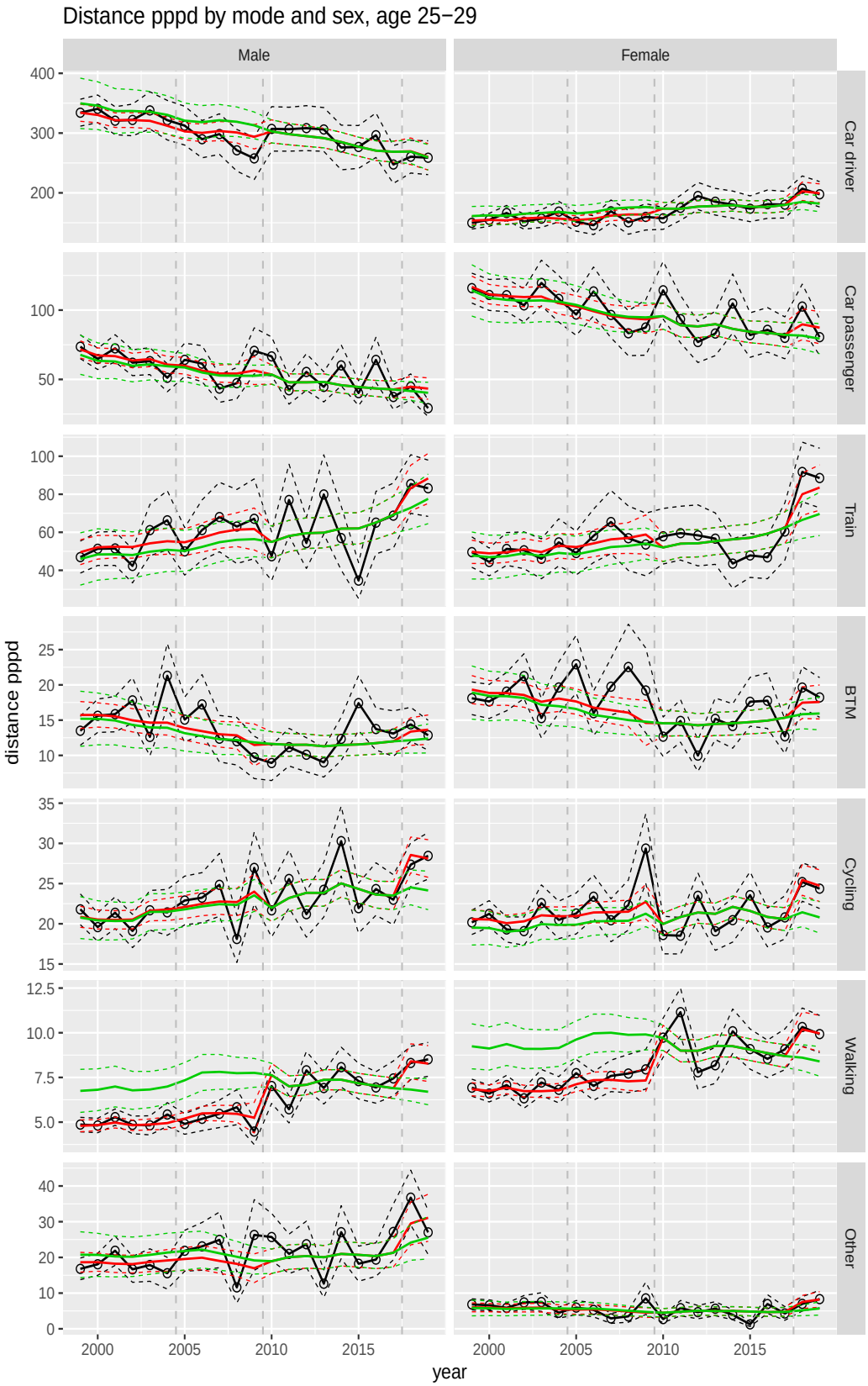

Figure A.159 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

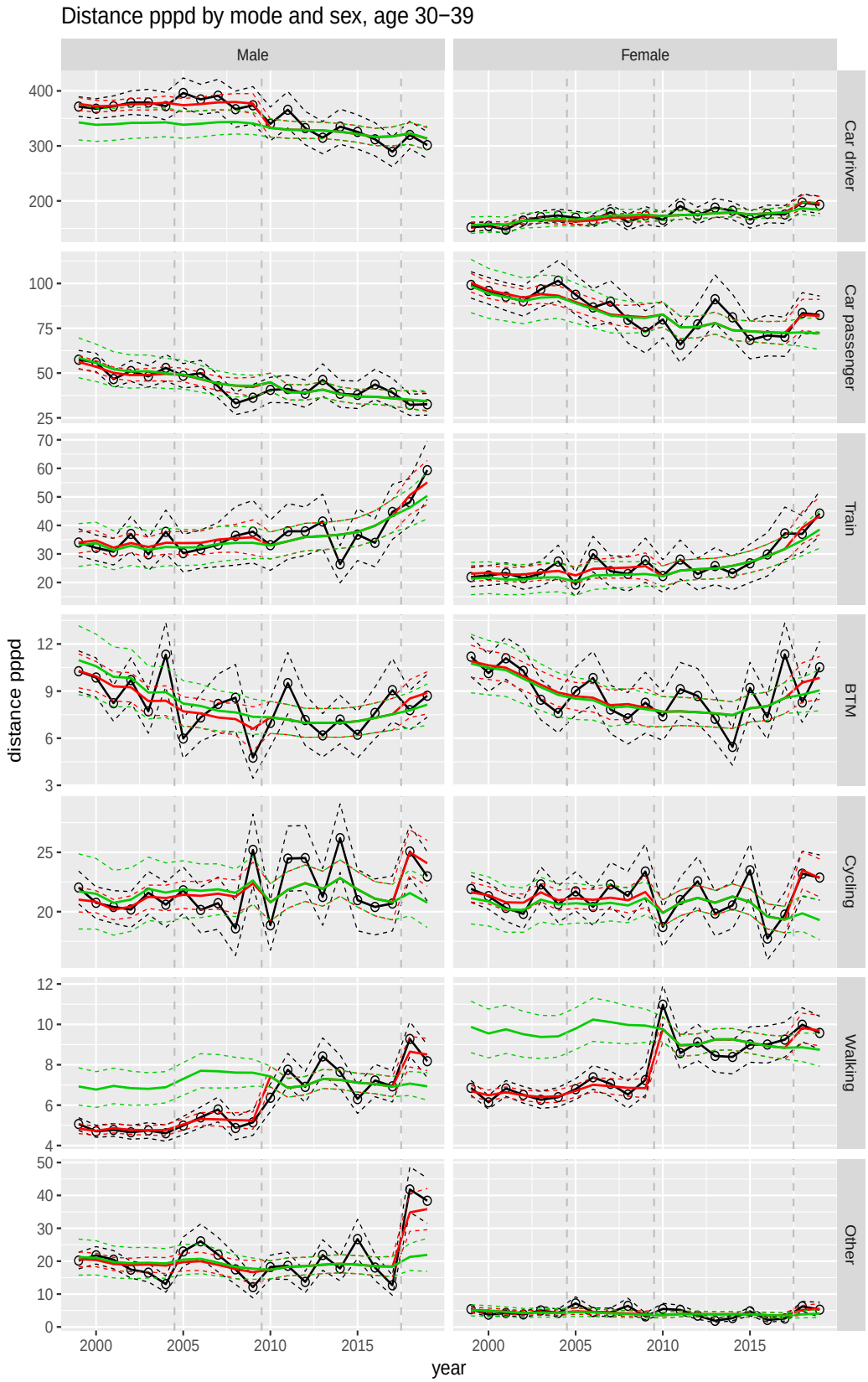

Figure A.160 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

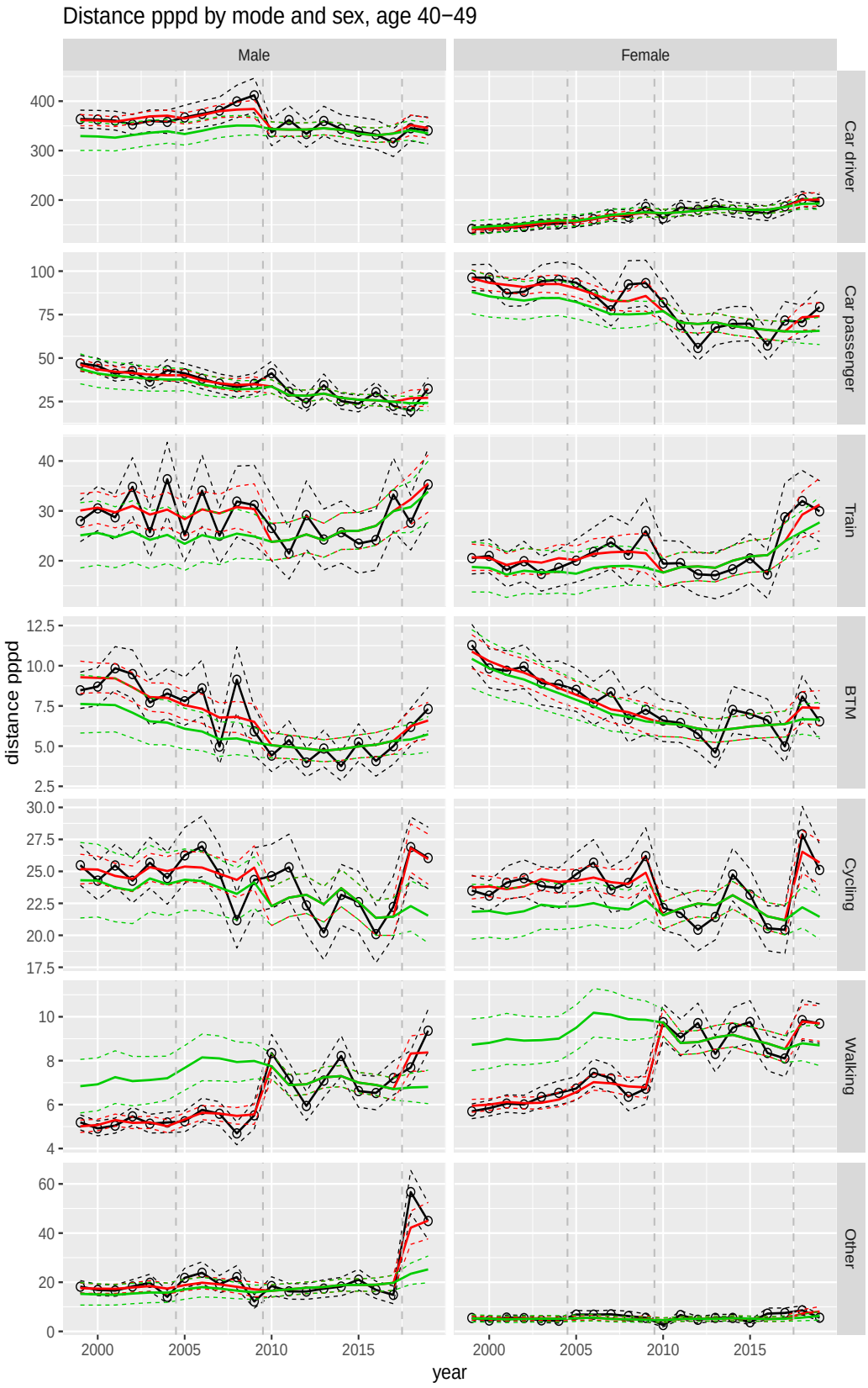

Figure A.161 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure A.162 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

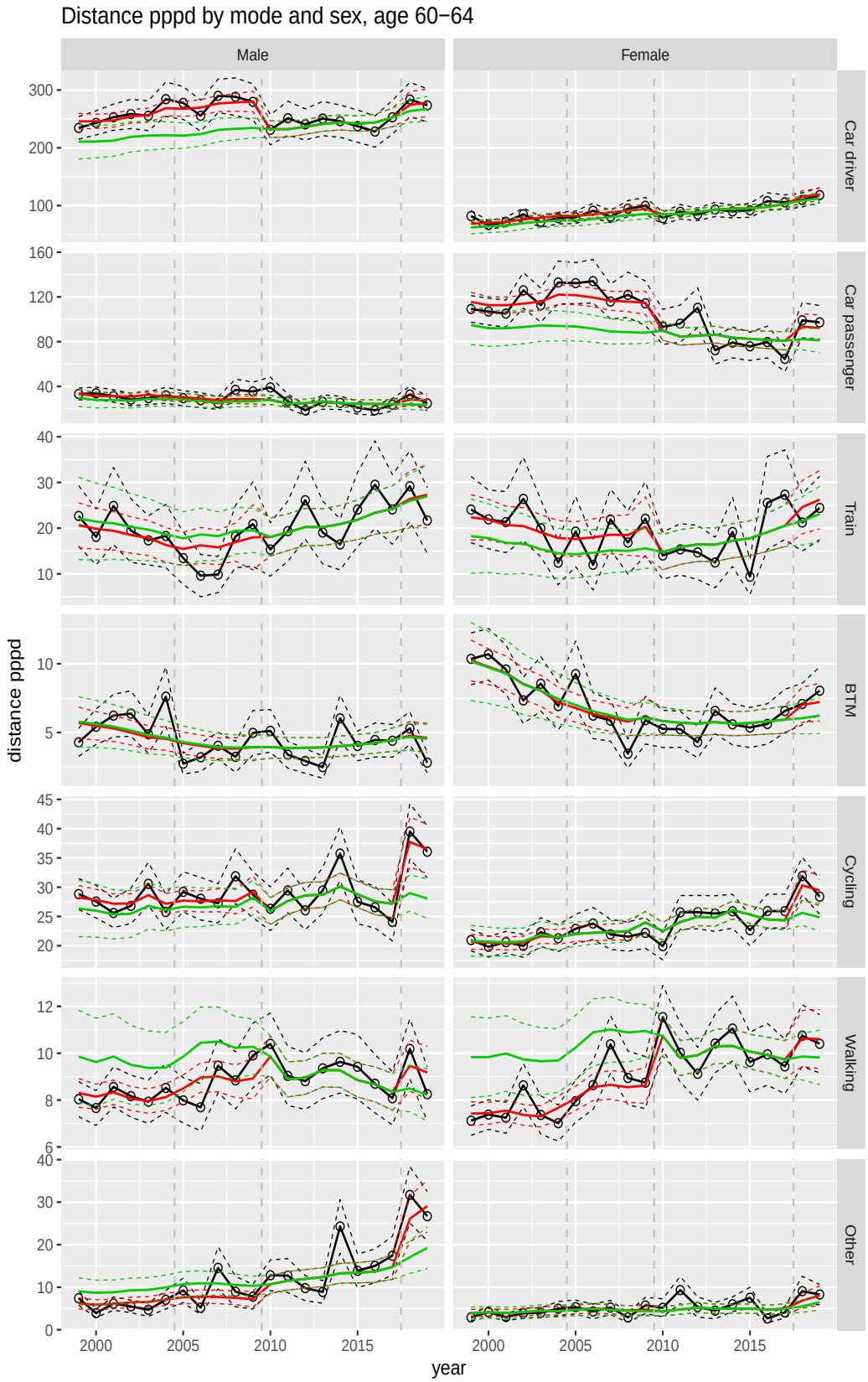

Figure A.163 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

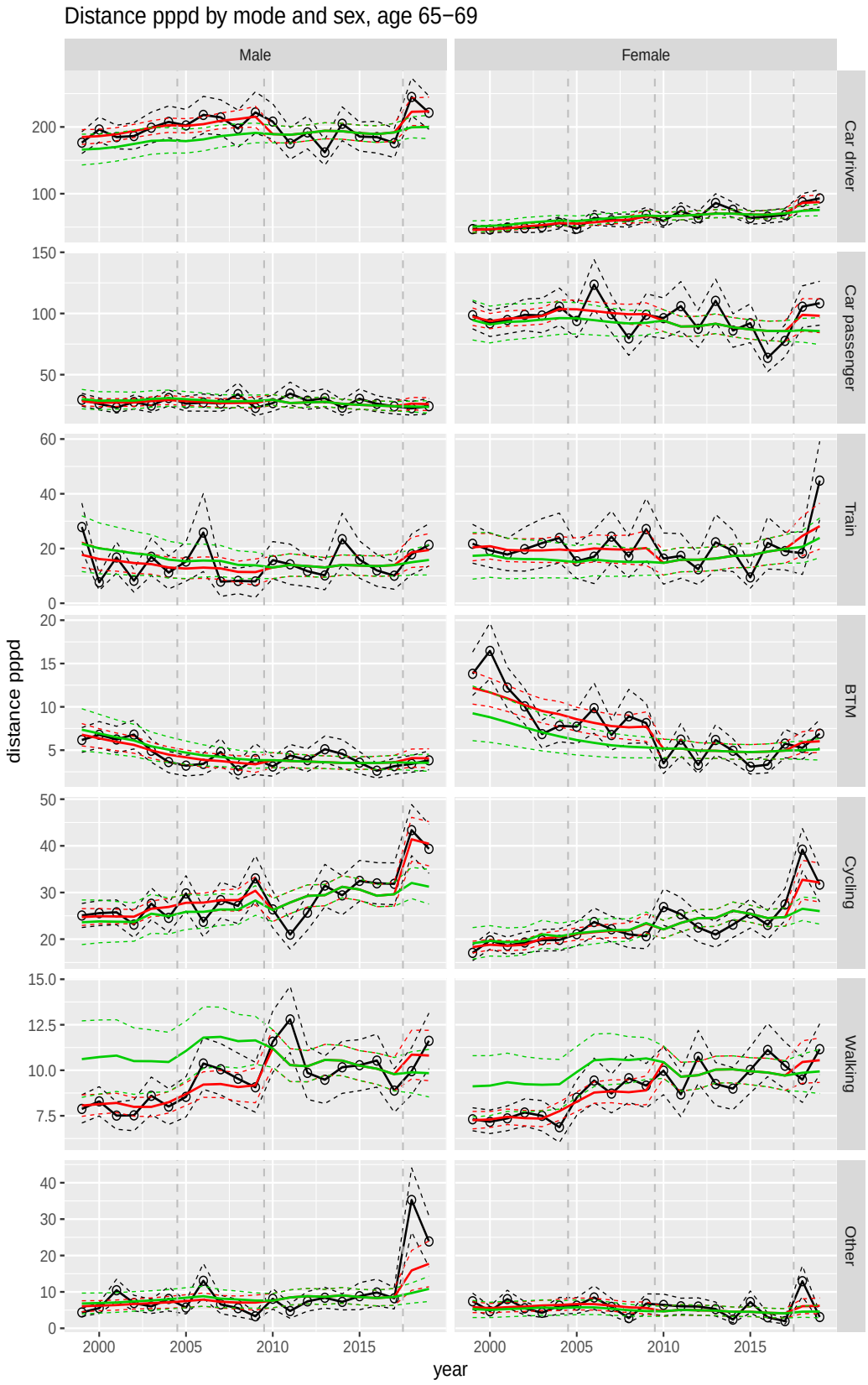

Figure A.164 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

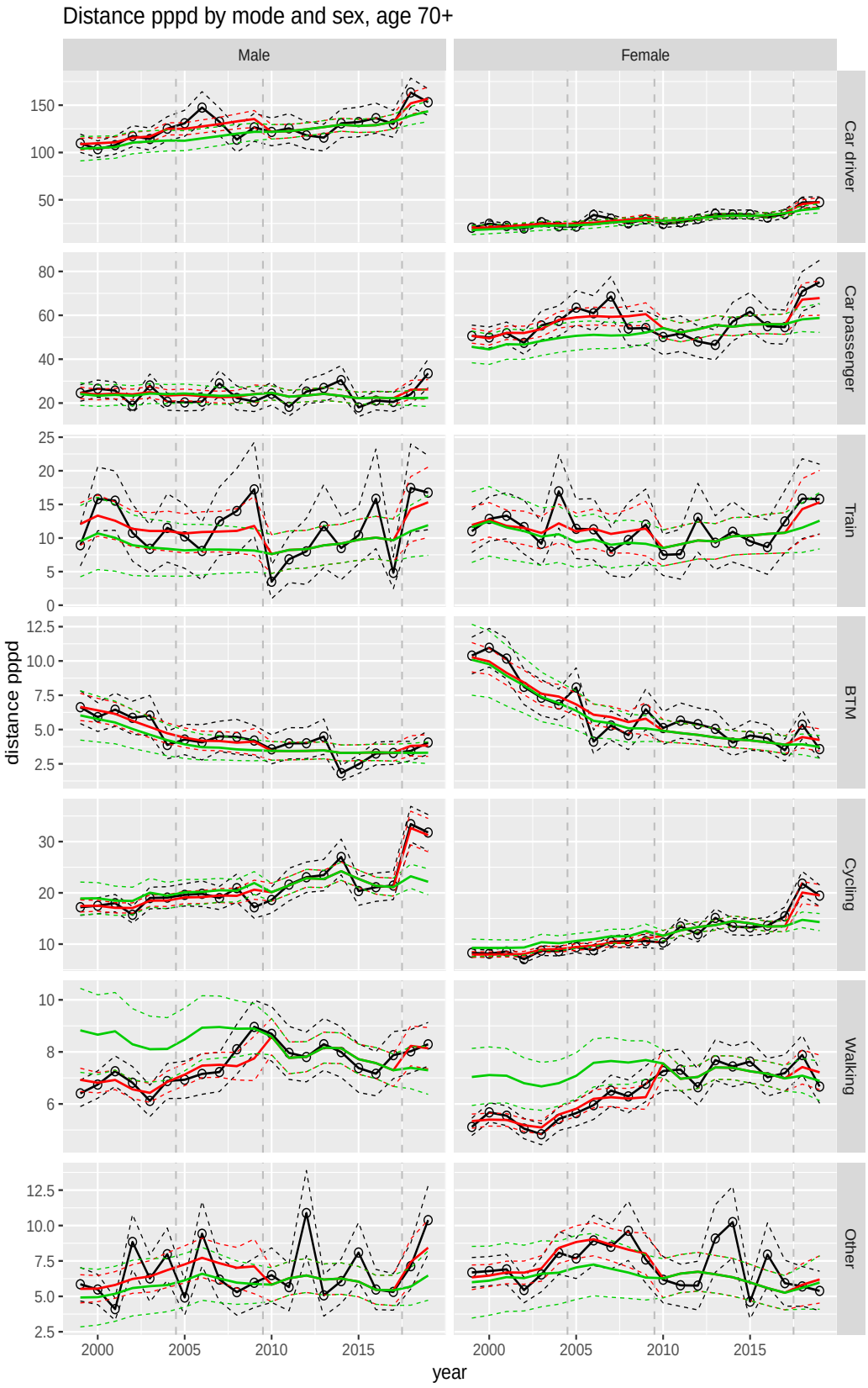

Figure A.165 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

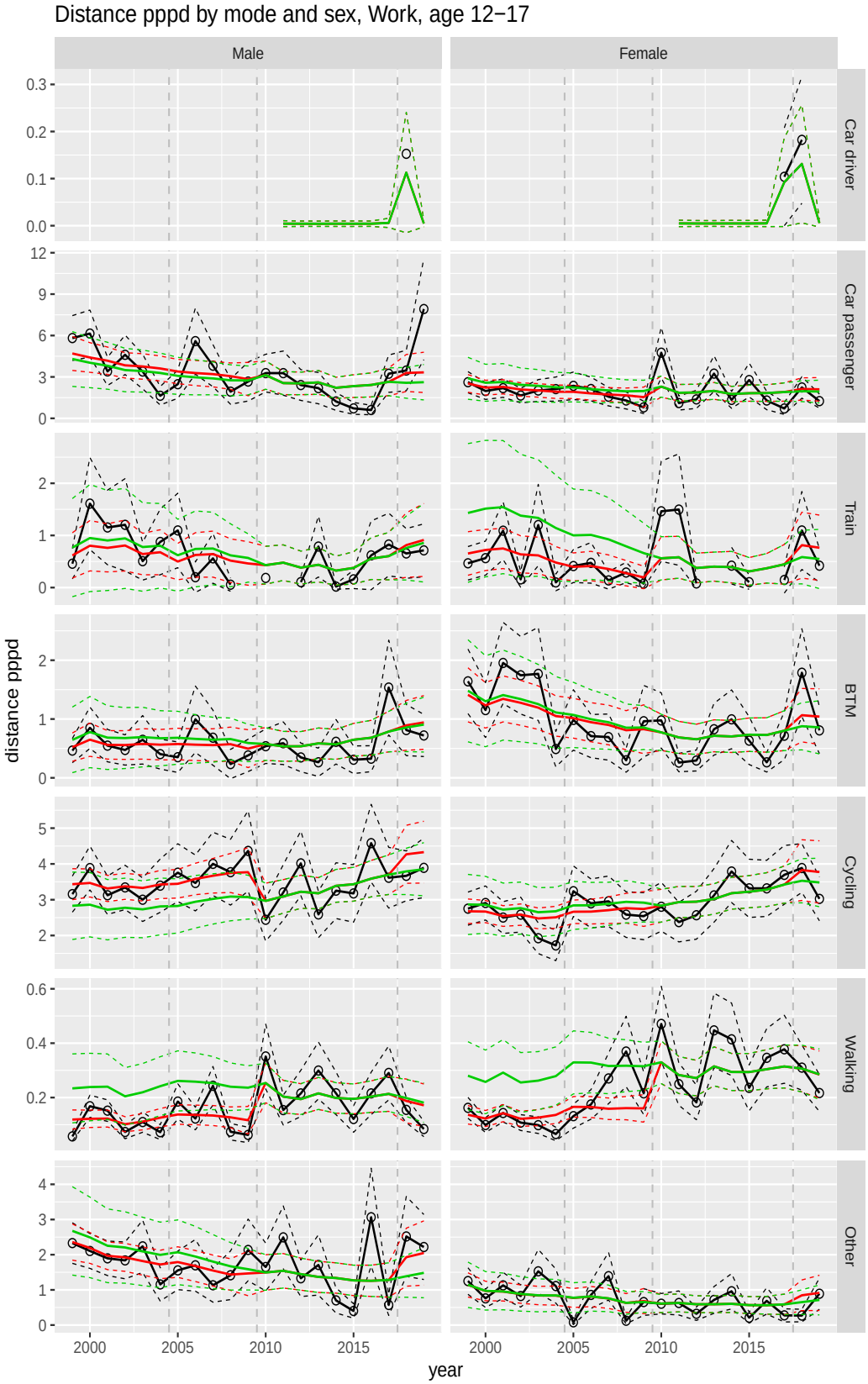

Figure A.166 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

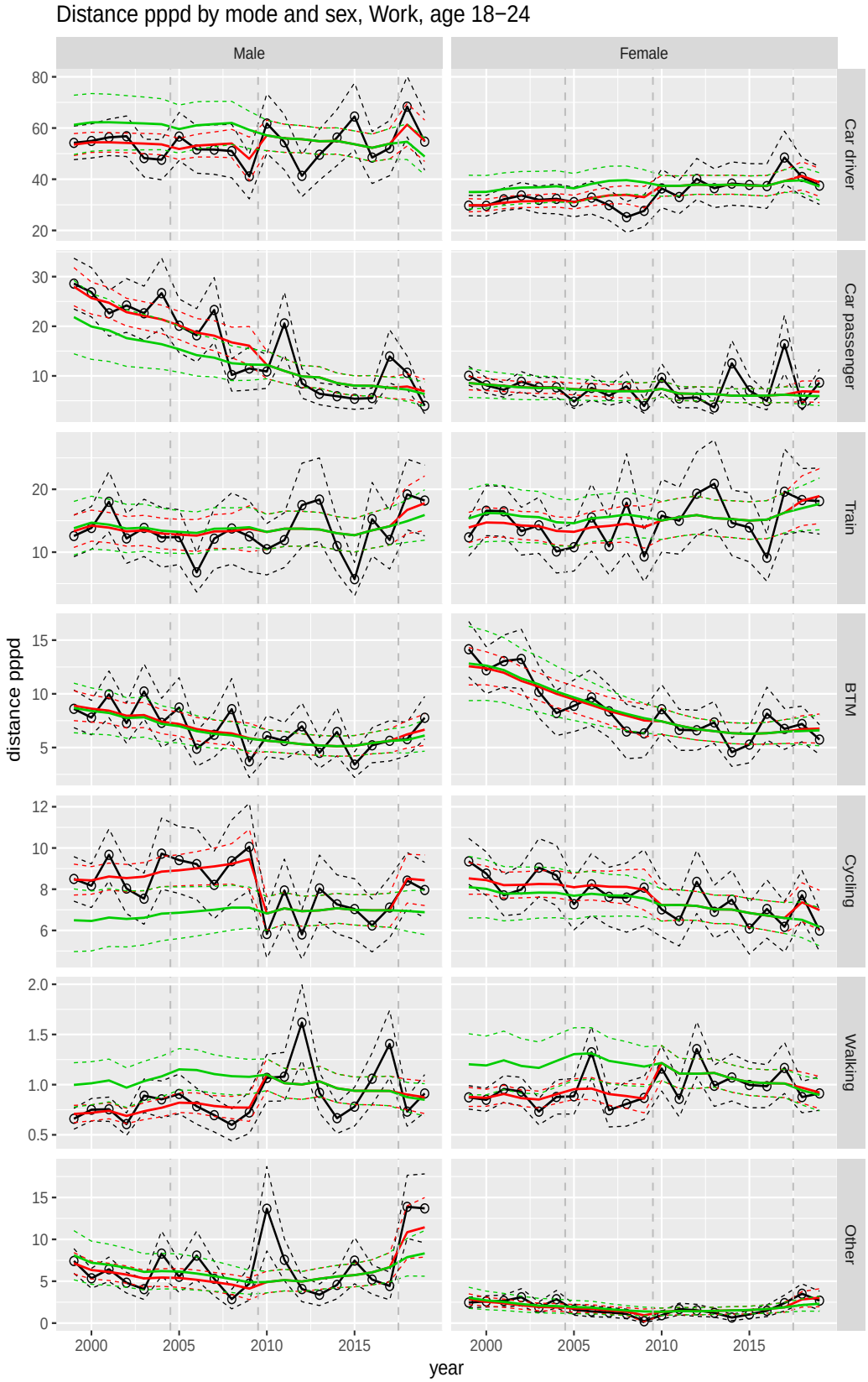

Figure A.167 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

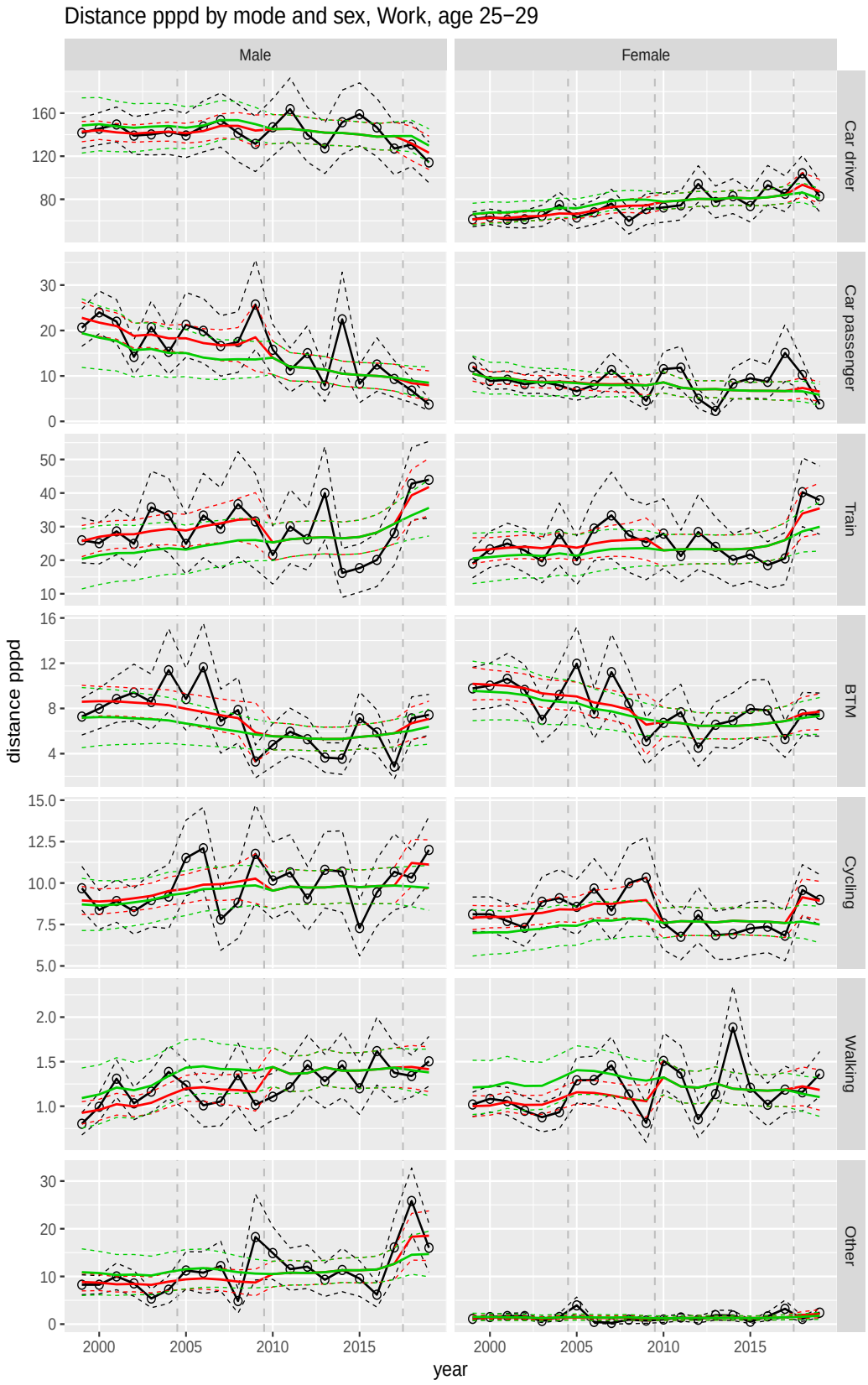

Figure A.168 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

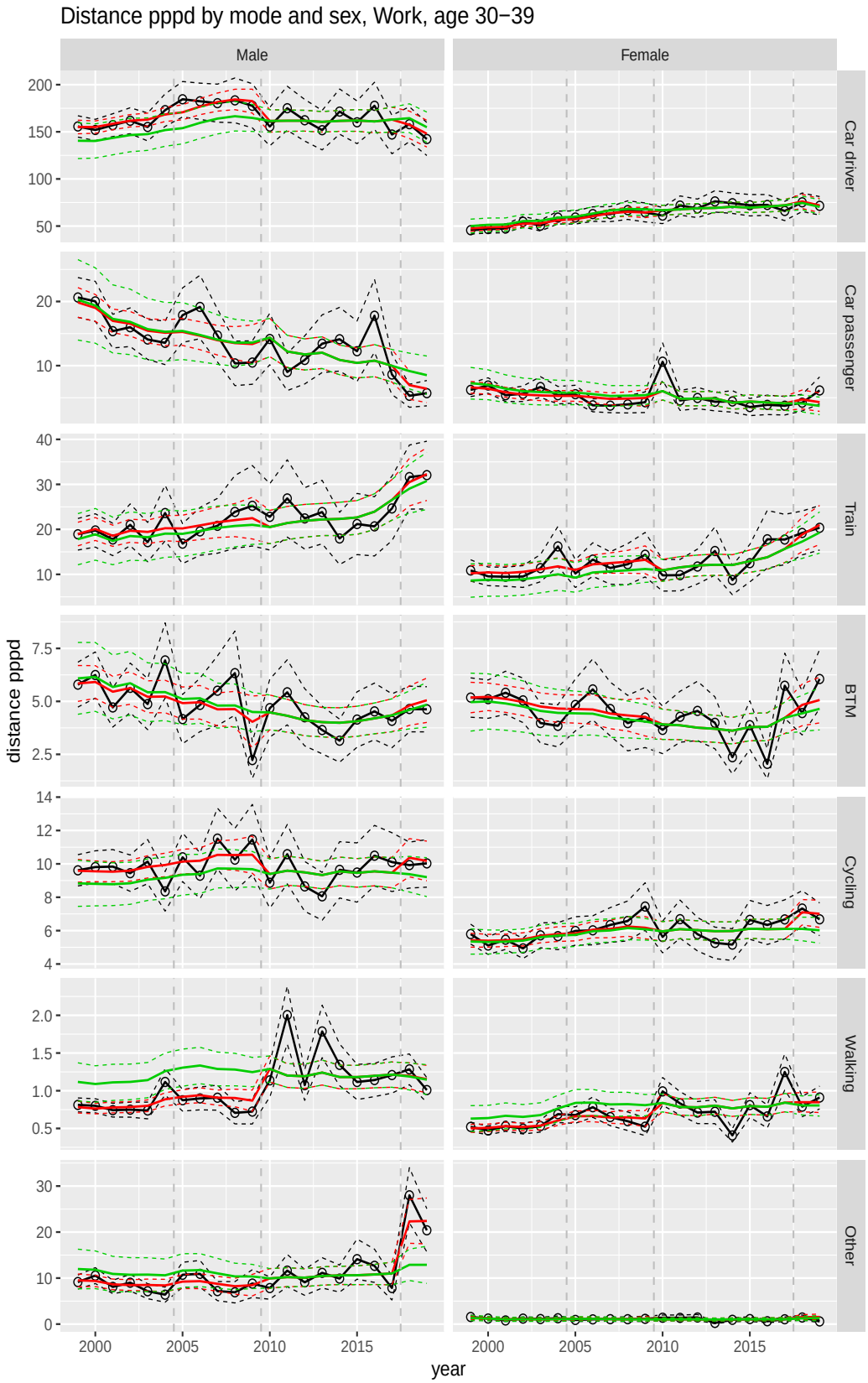

Figure A.169 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

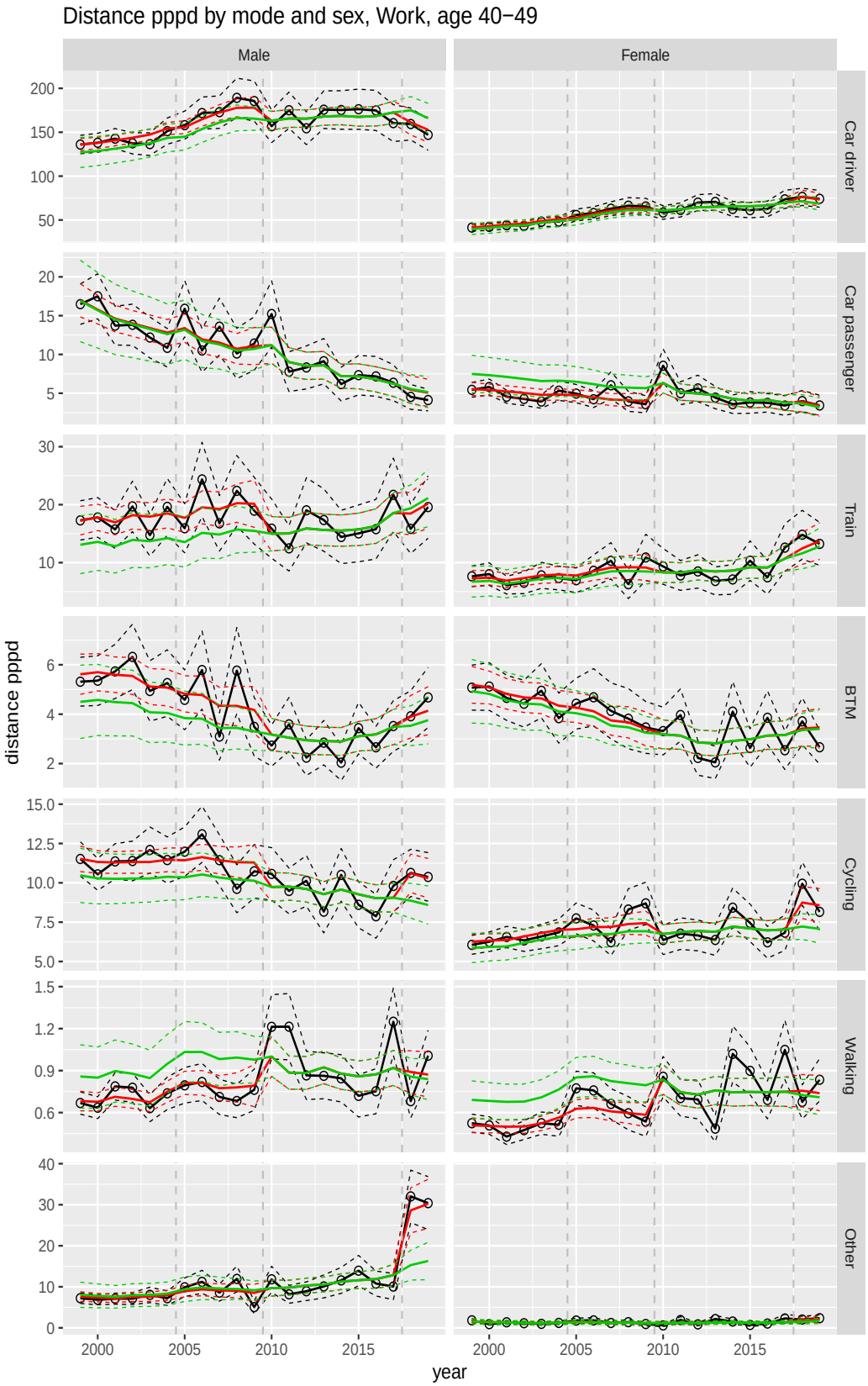

Figure A.170 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

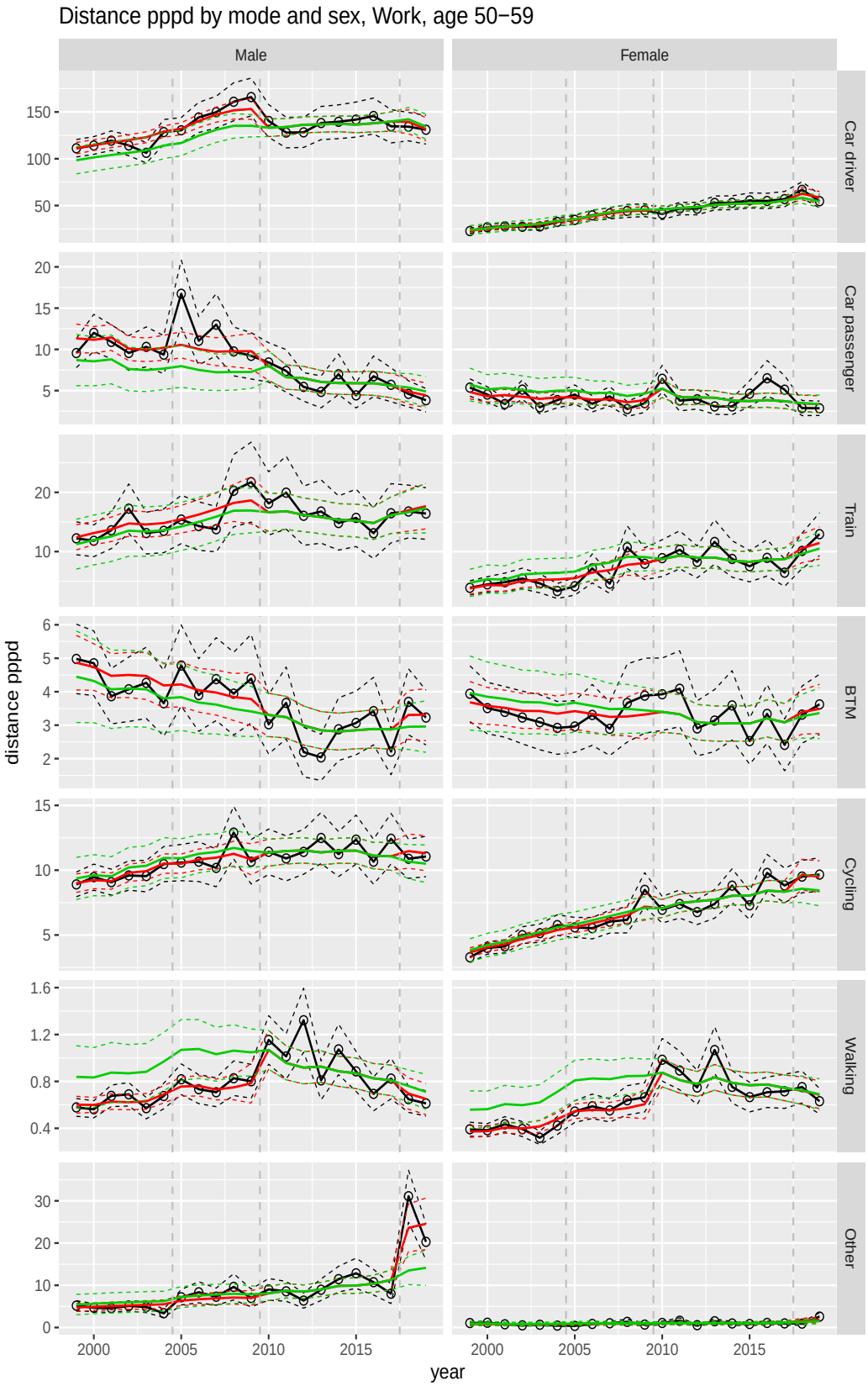

Figure A.171 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

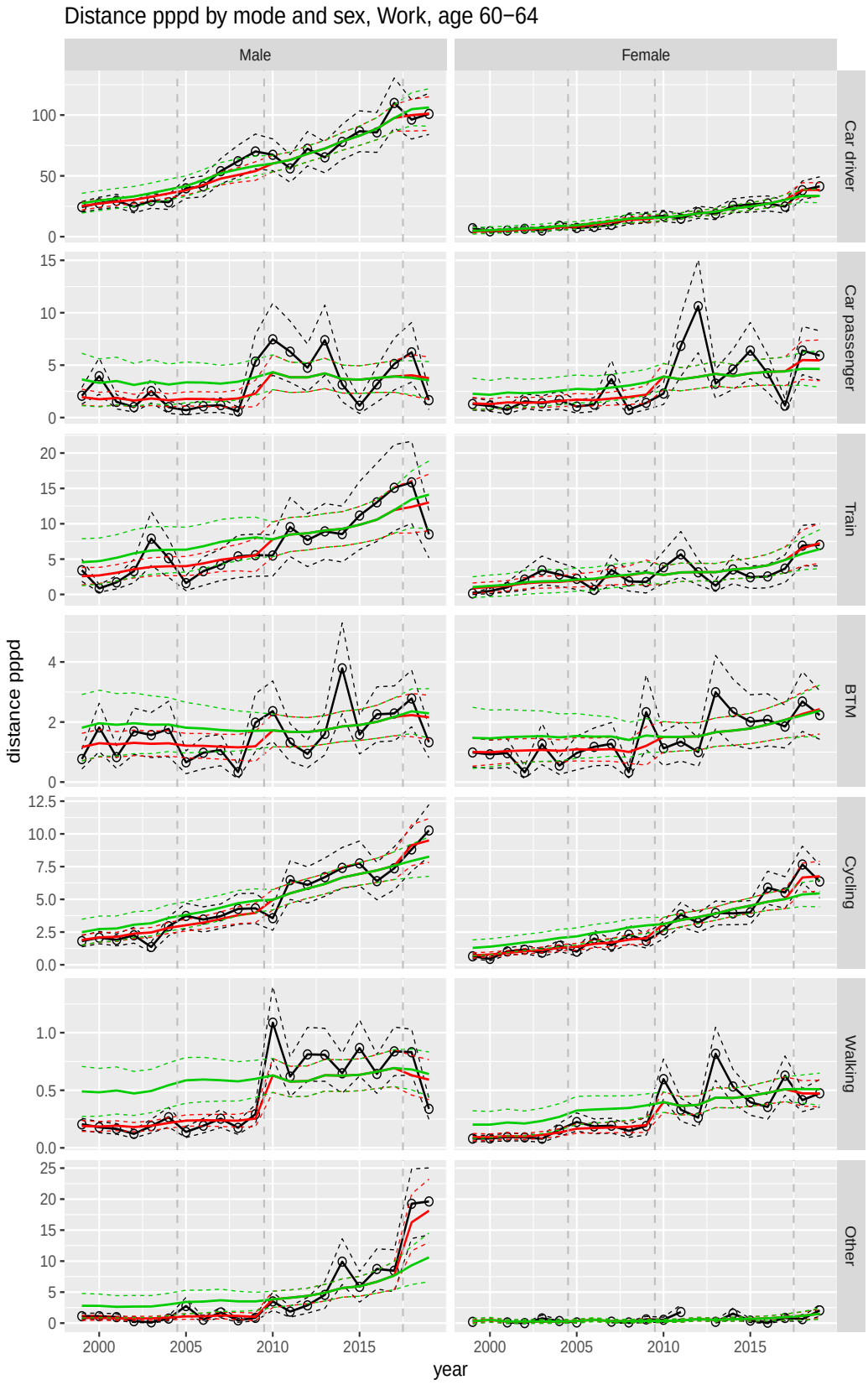

Figure A.172 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

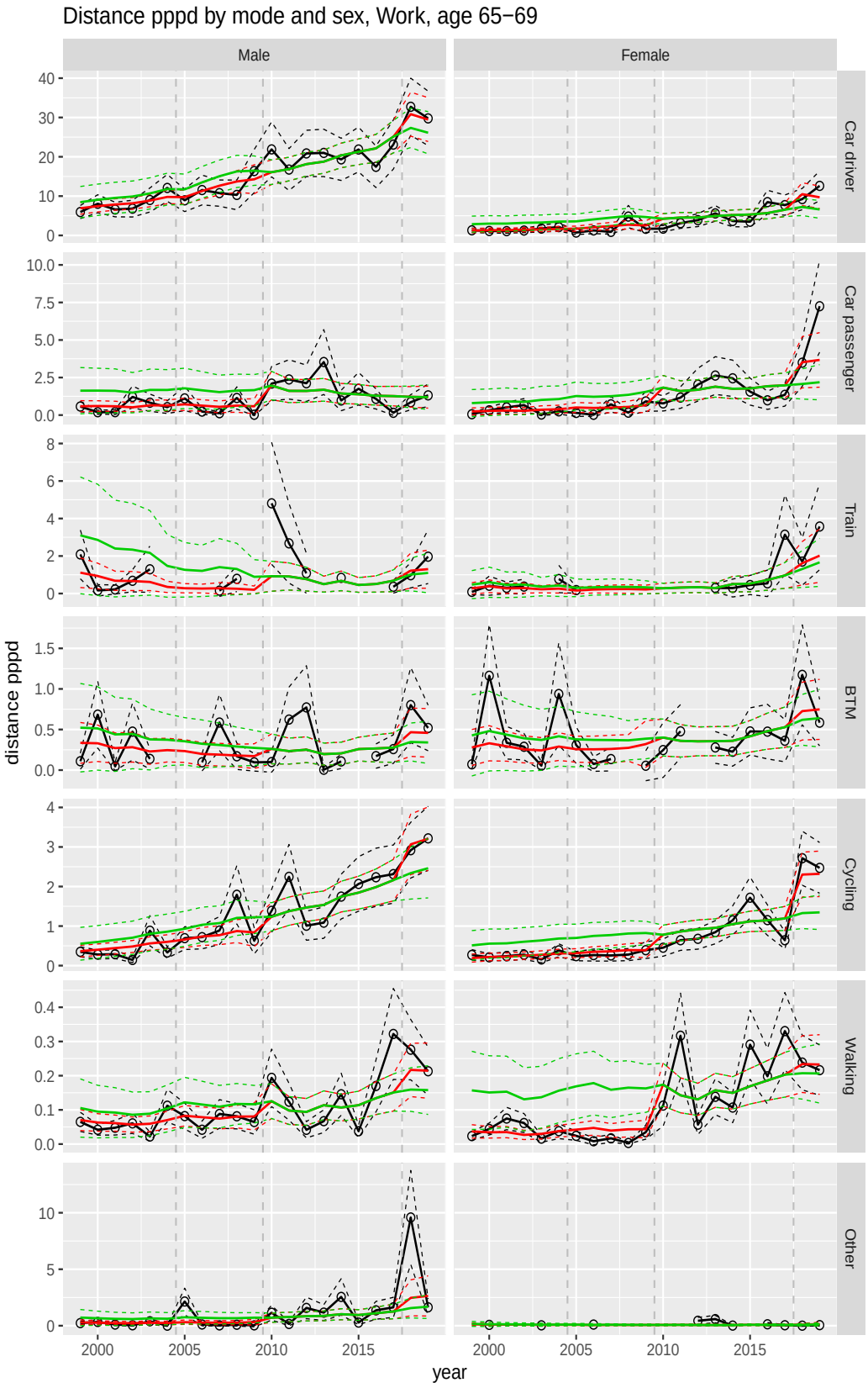

Figure A.173 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

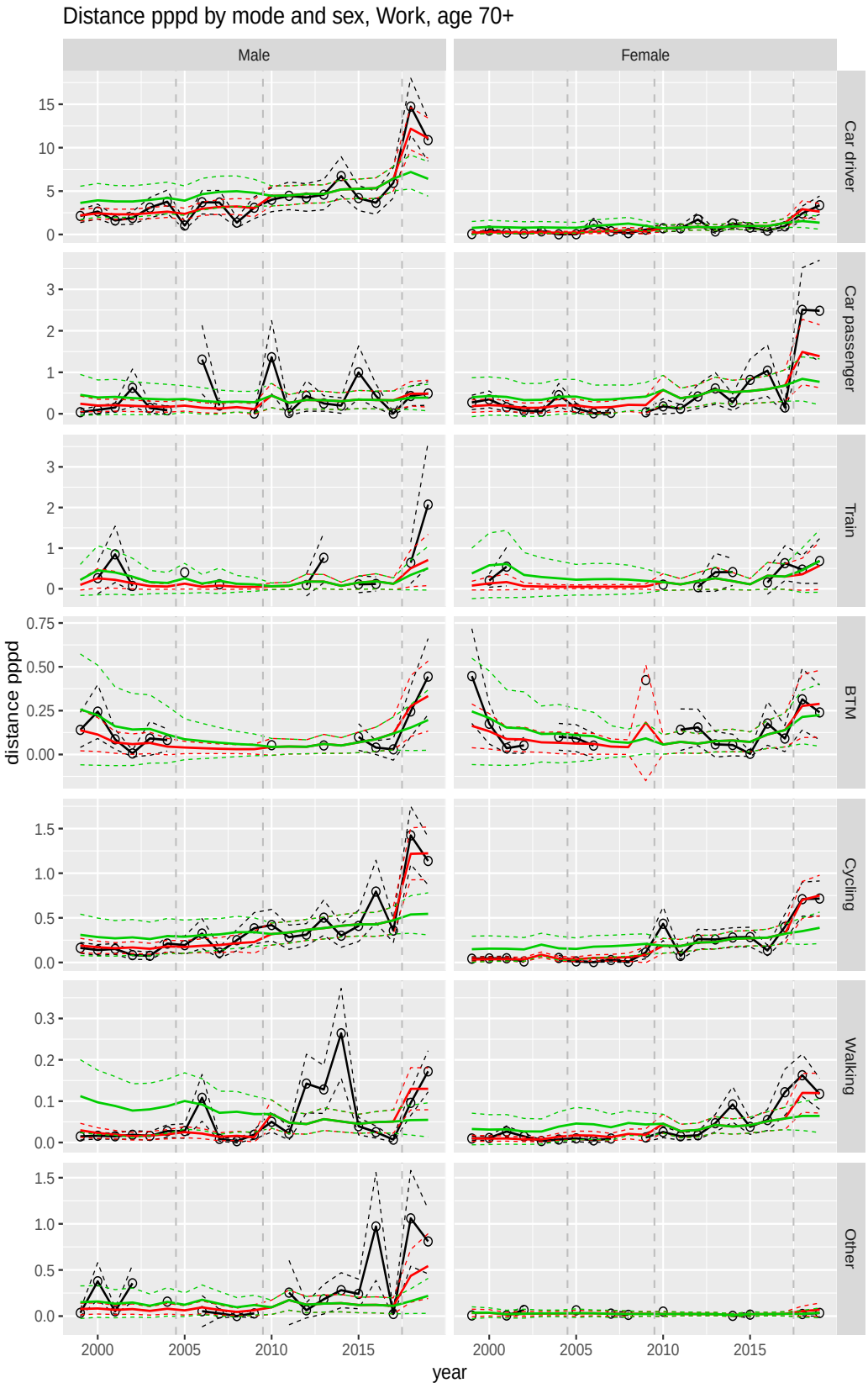

Figure A.174 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

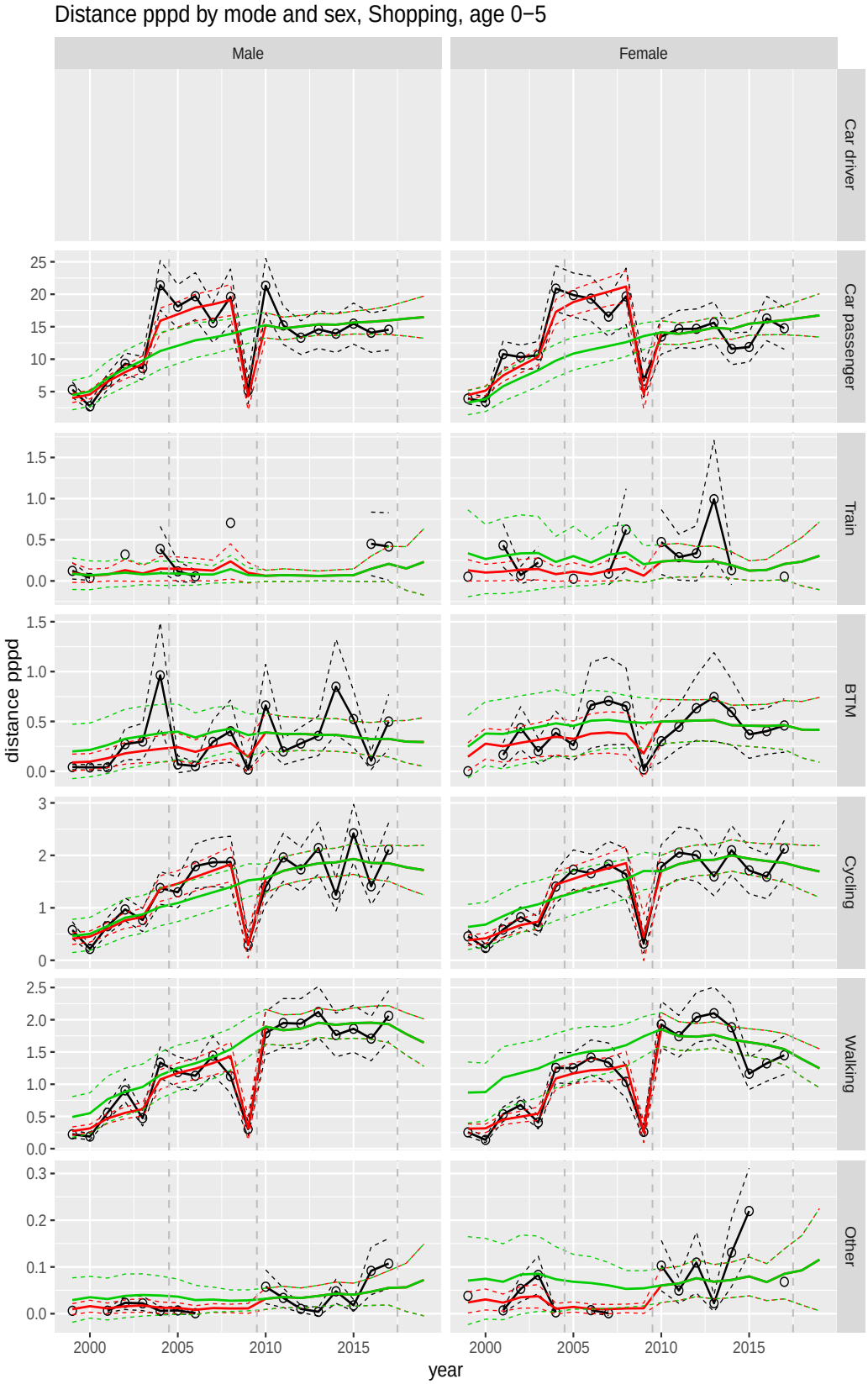

Figure A.175 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

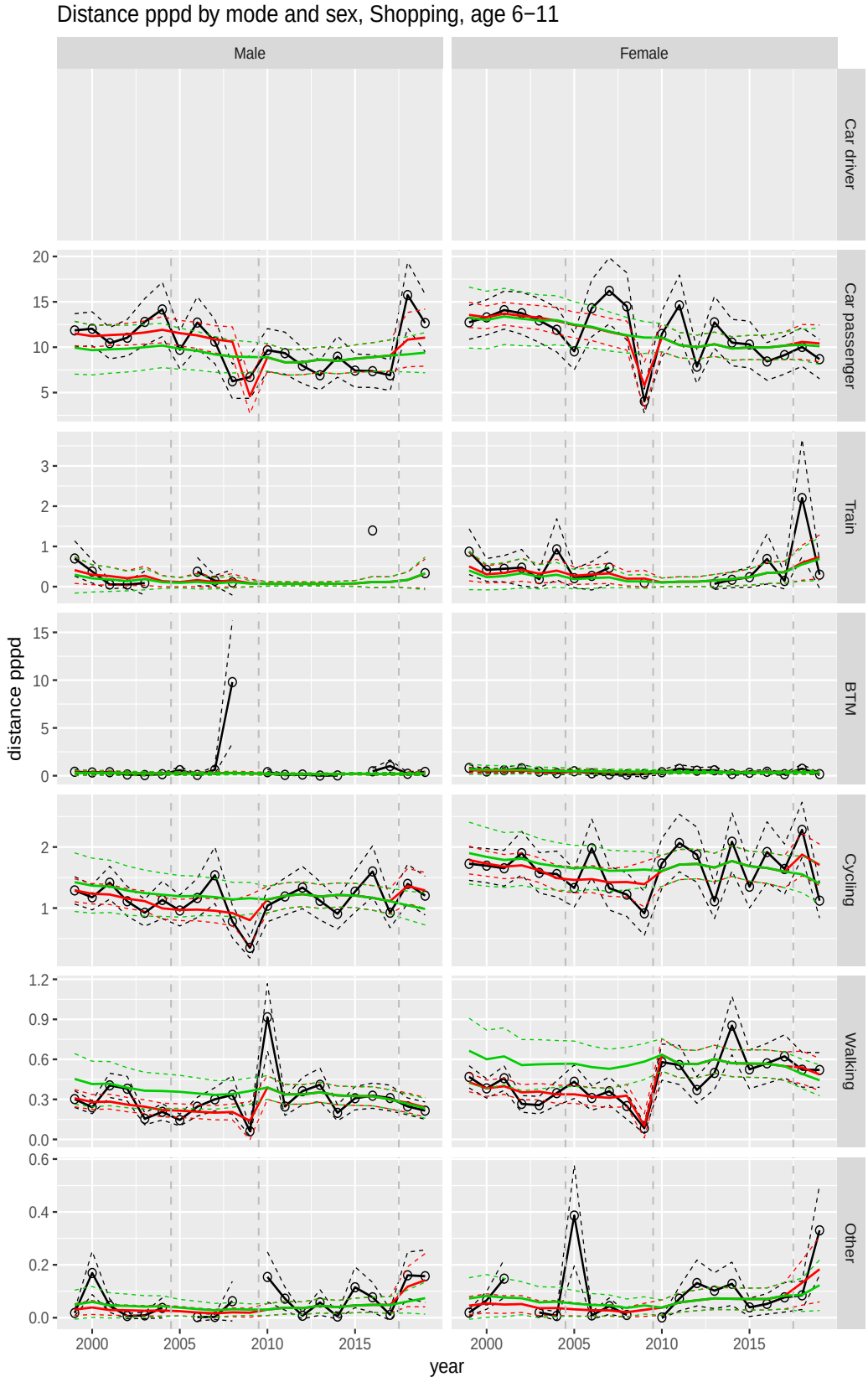

Figure A.176 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

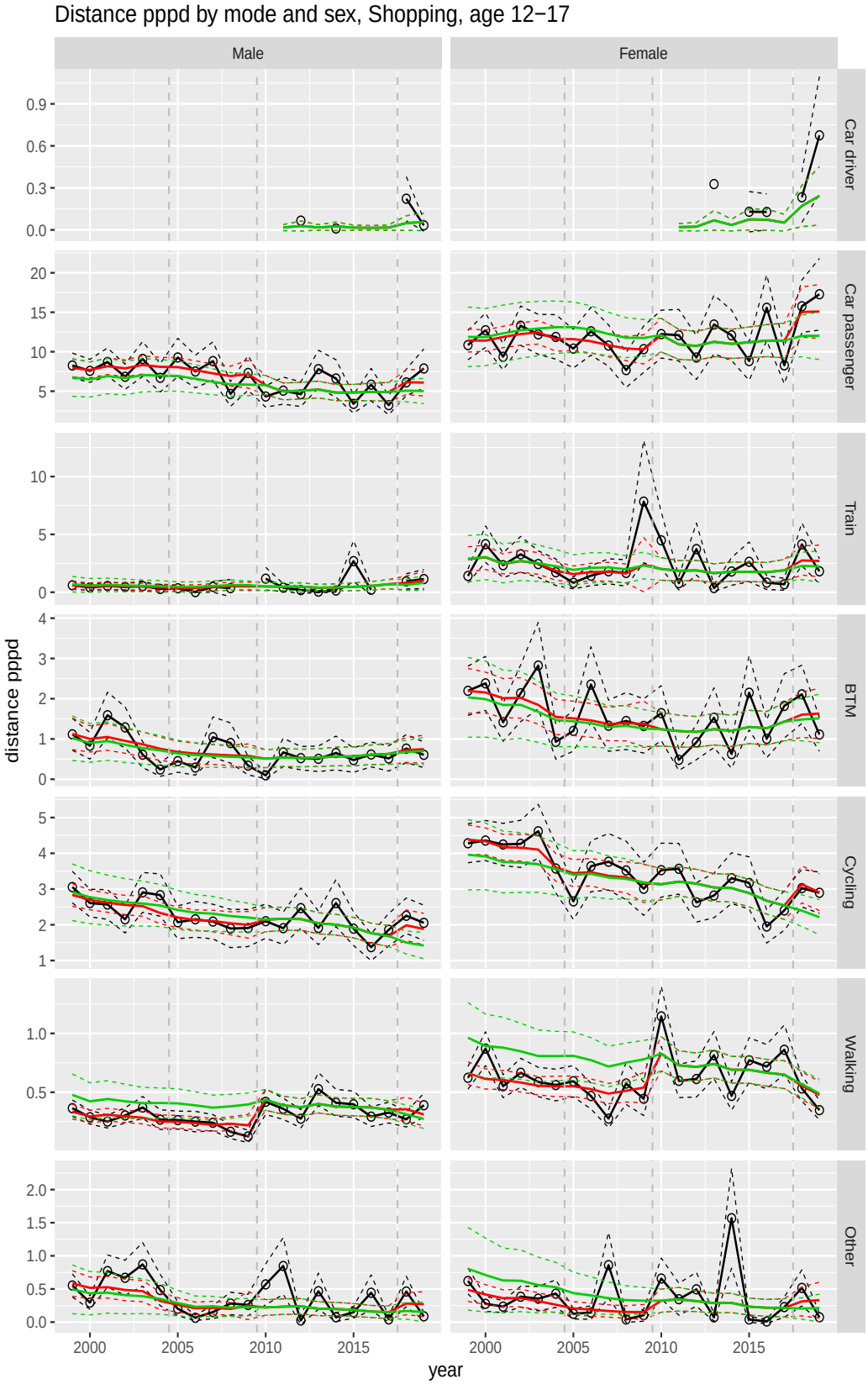

Figure A.177 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

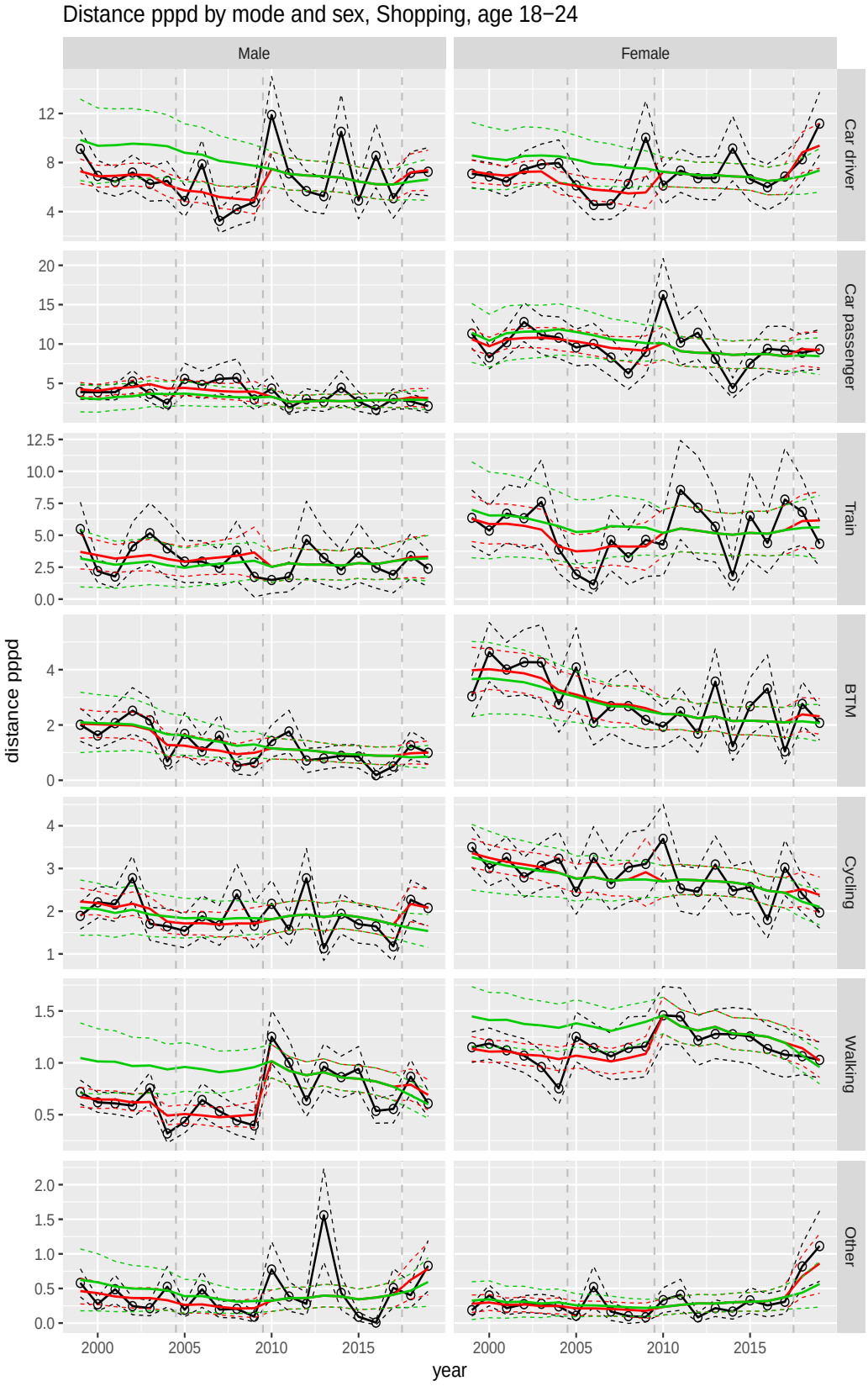

Figure A.178 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

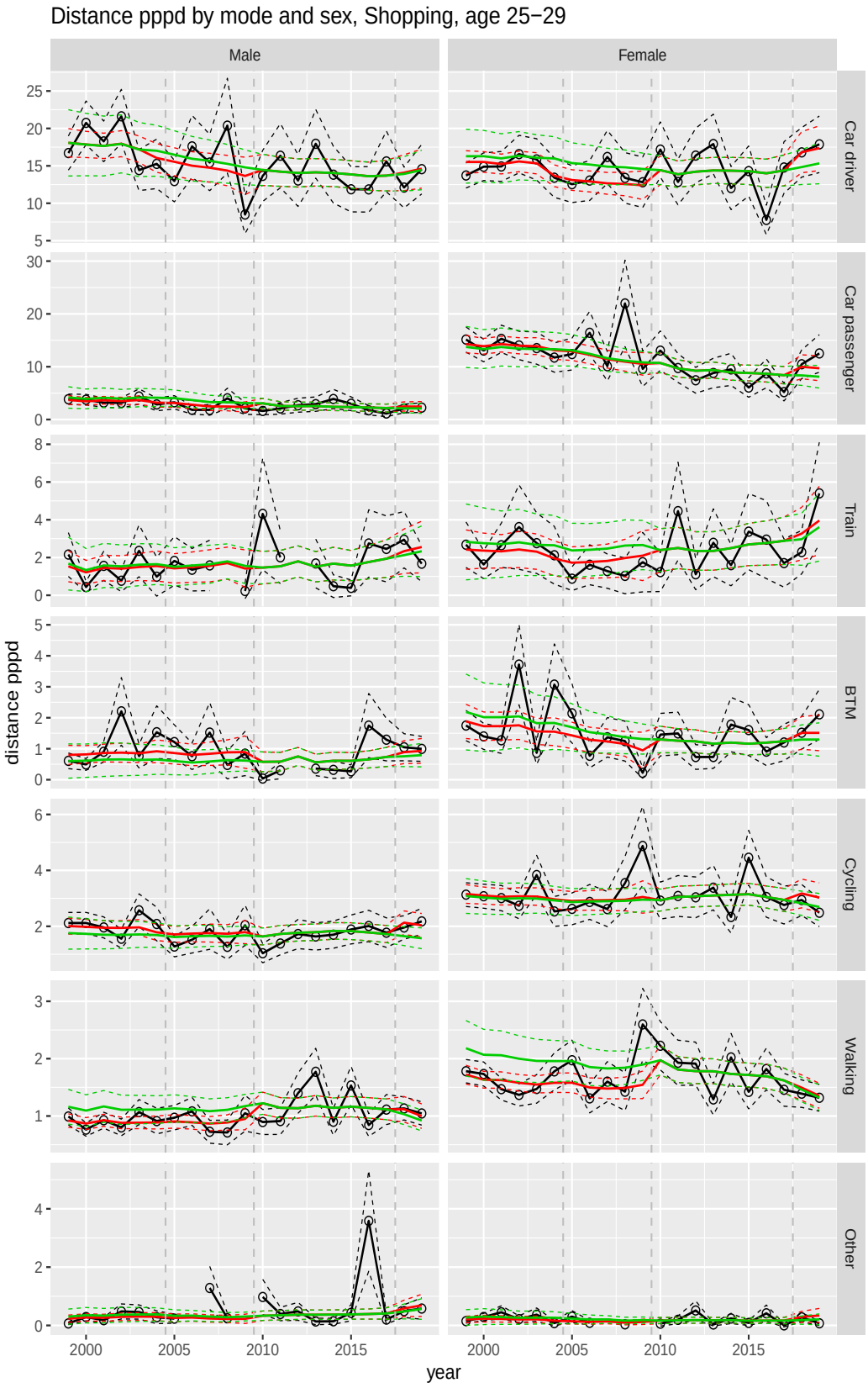

Figure A.179 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

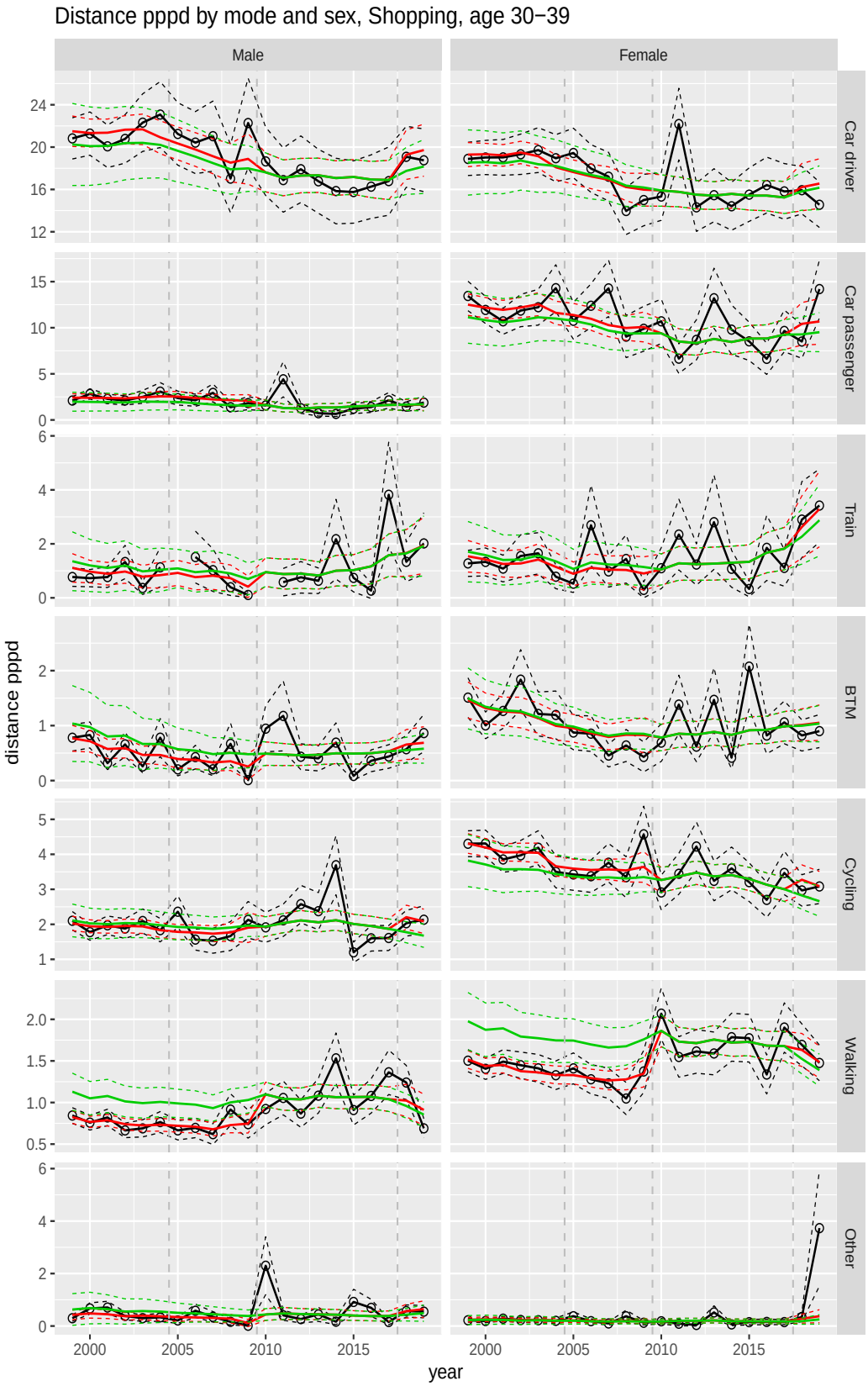

Figure A.180 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

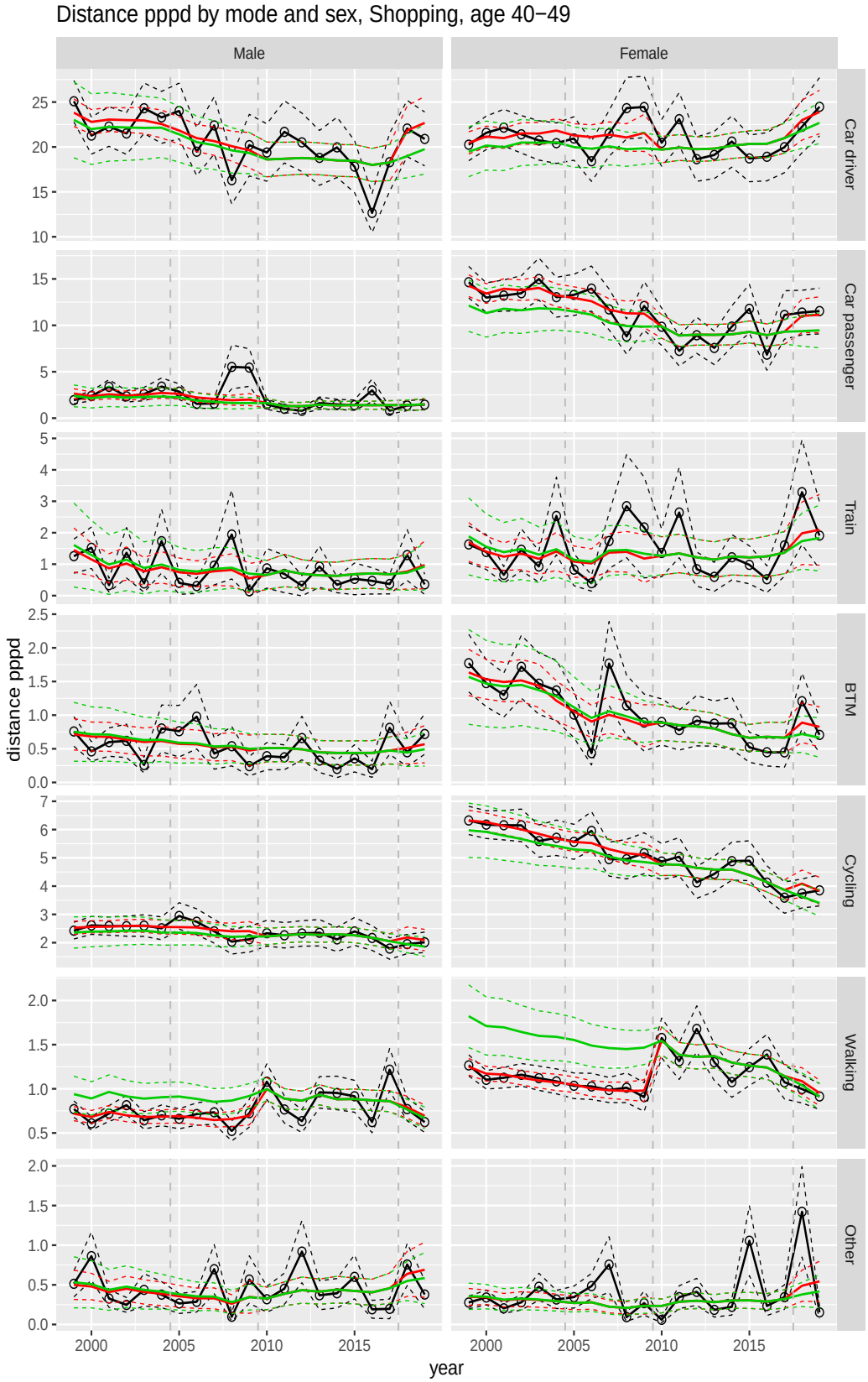

Figure A.181 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

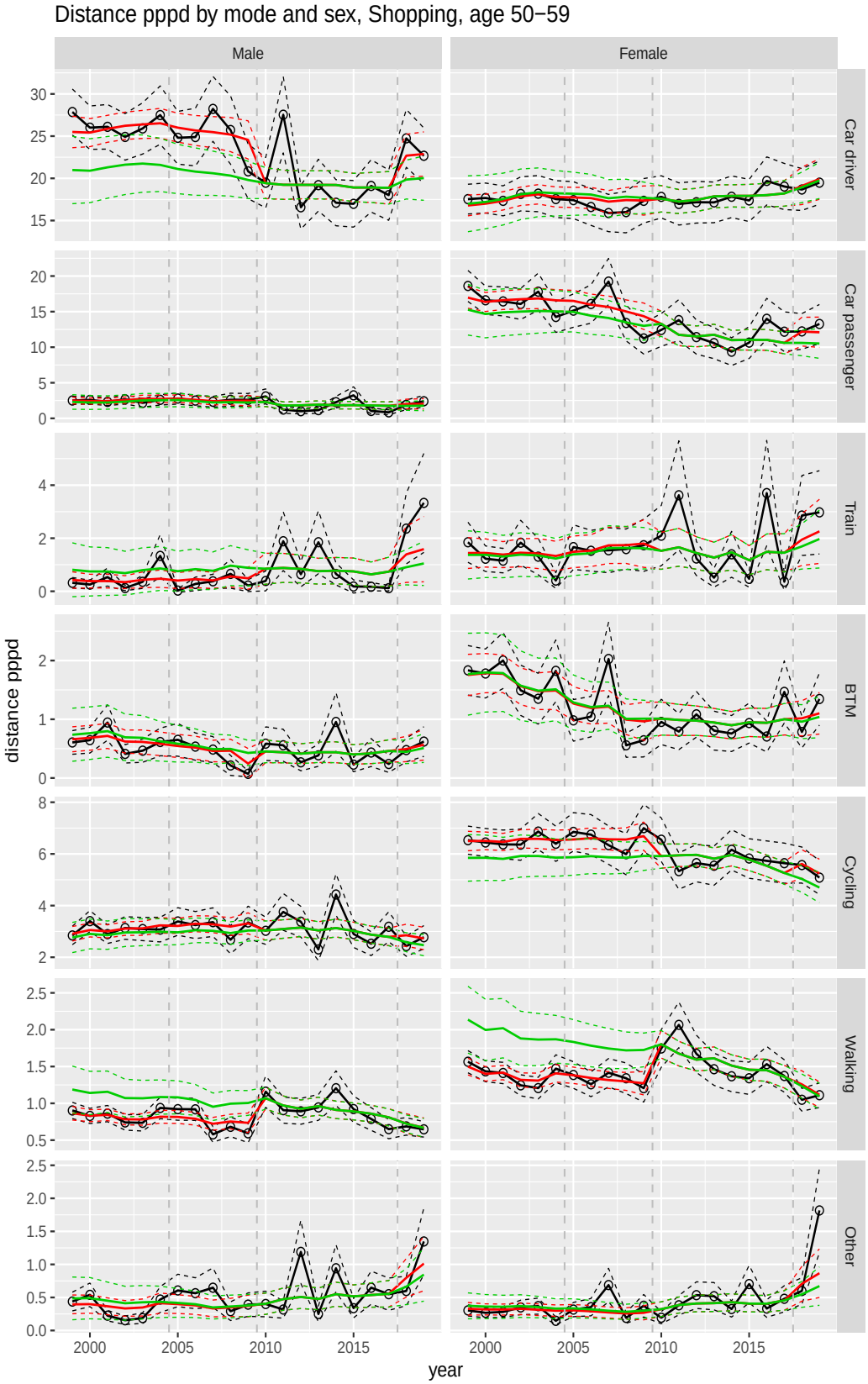

Figure A.182 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

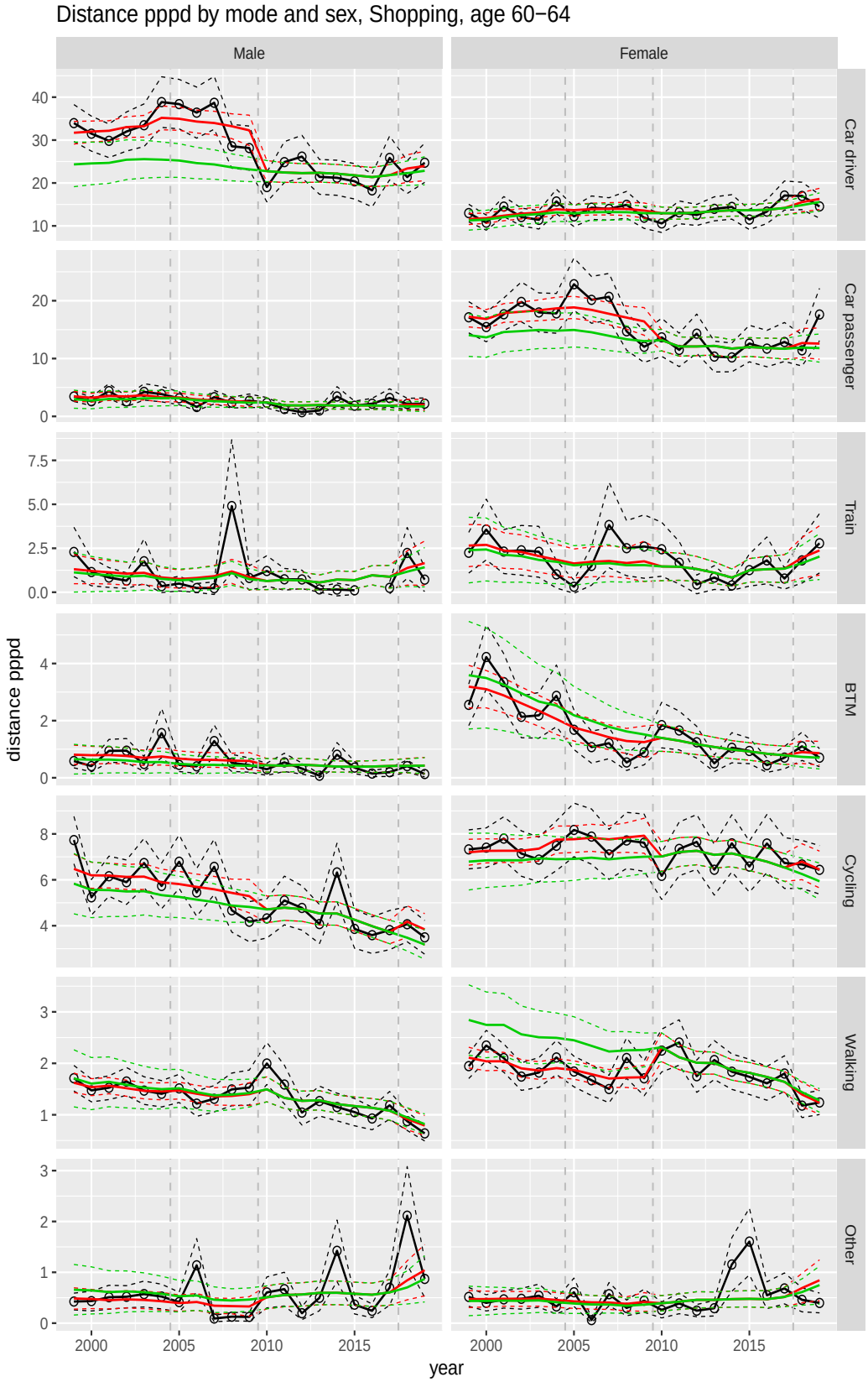

Figure A.183 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

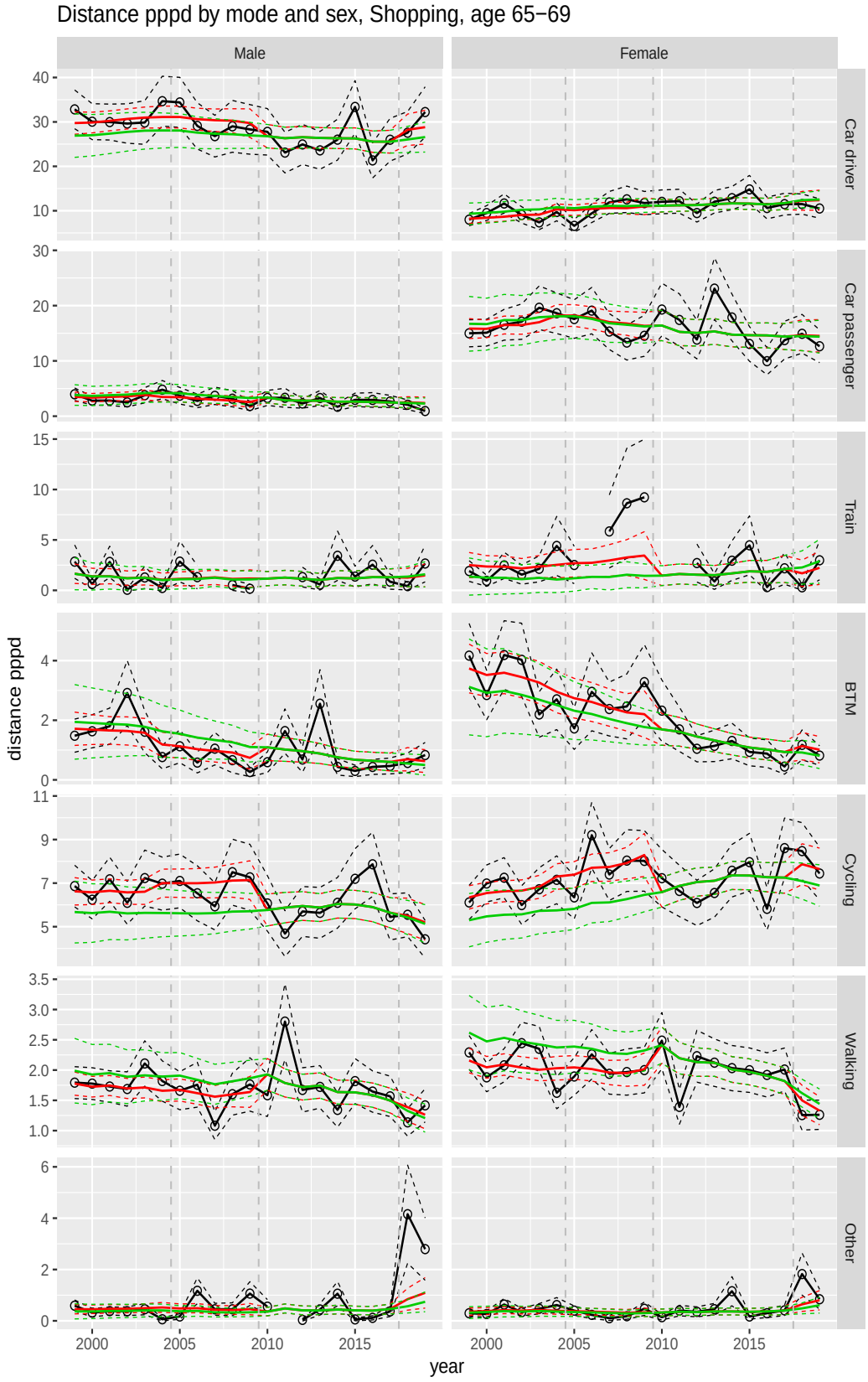

Figure A.184 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

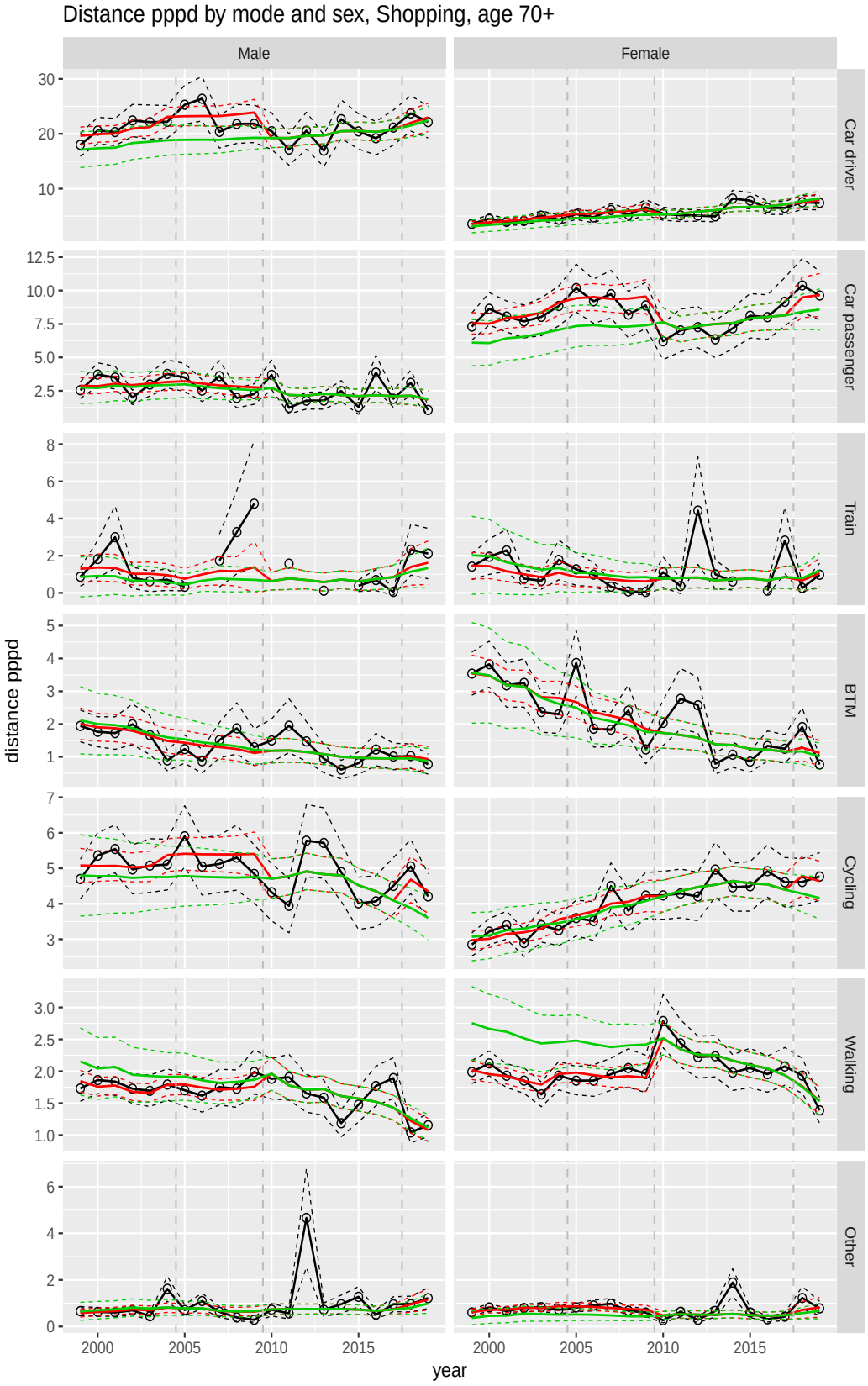

Figure A.185 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

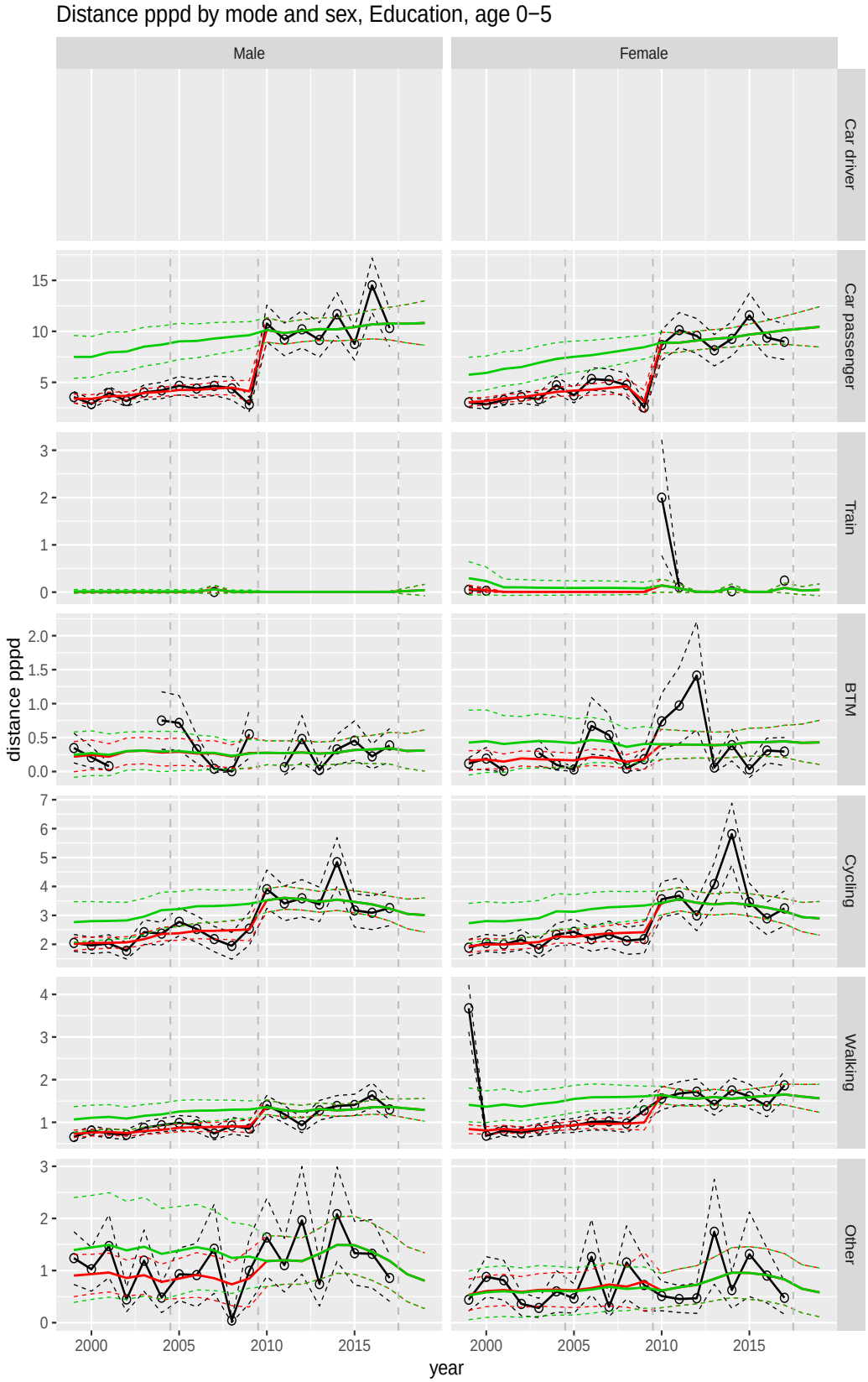

Figure A.186 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

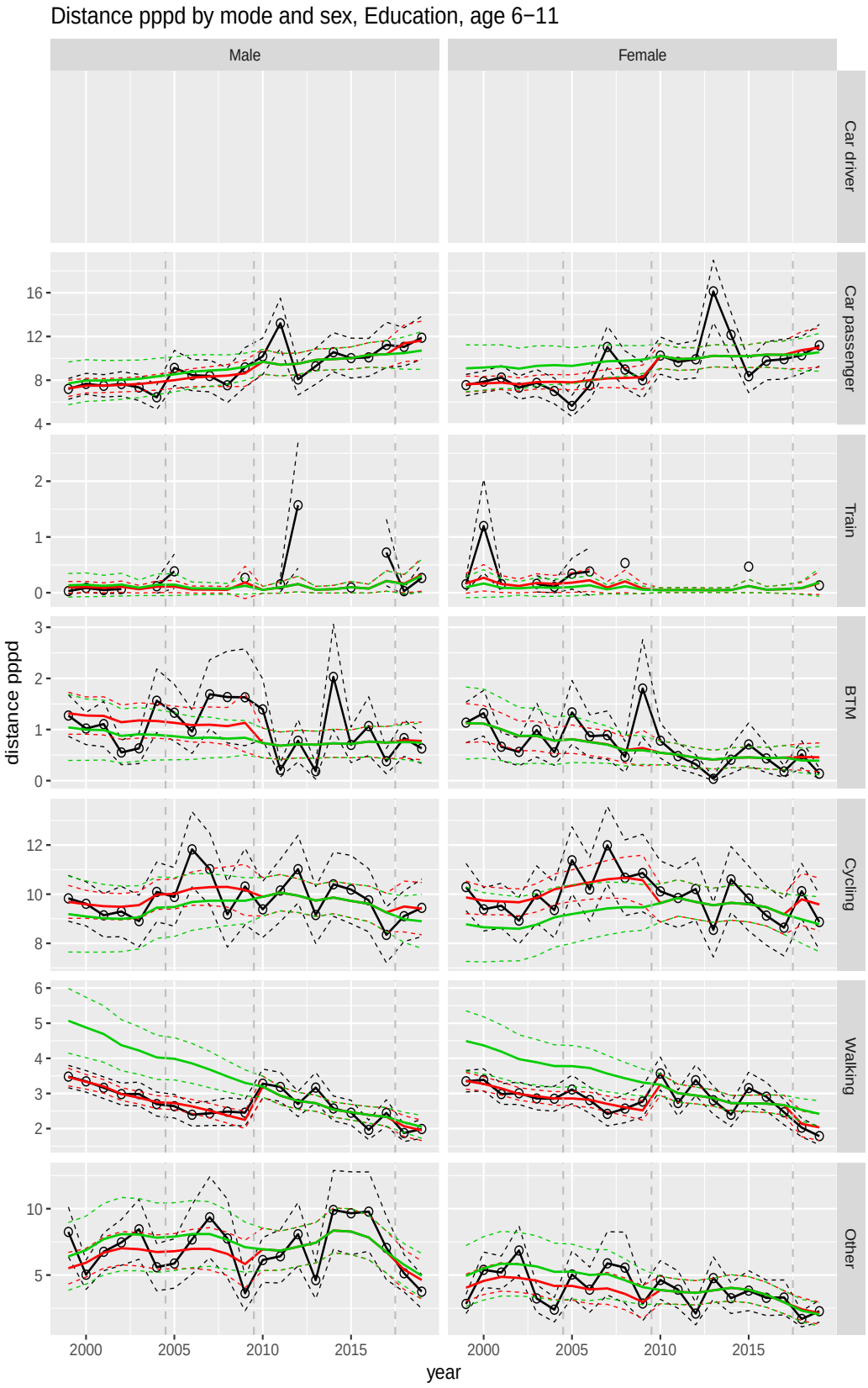

Figure A.187 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

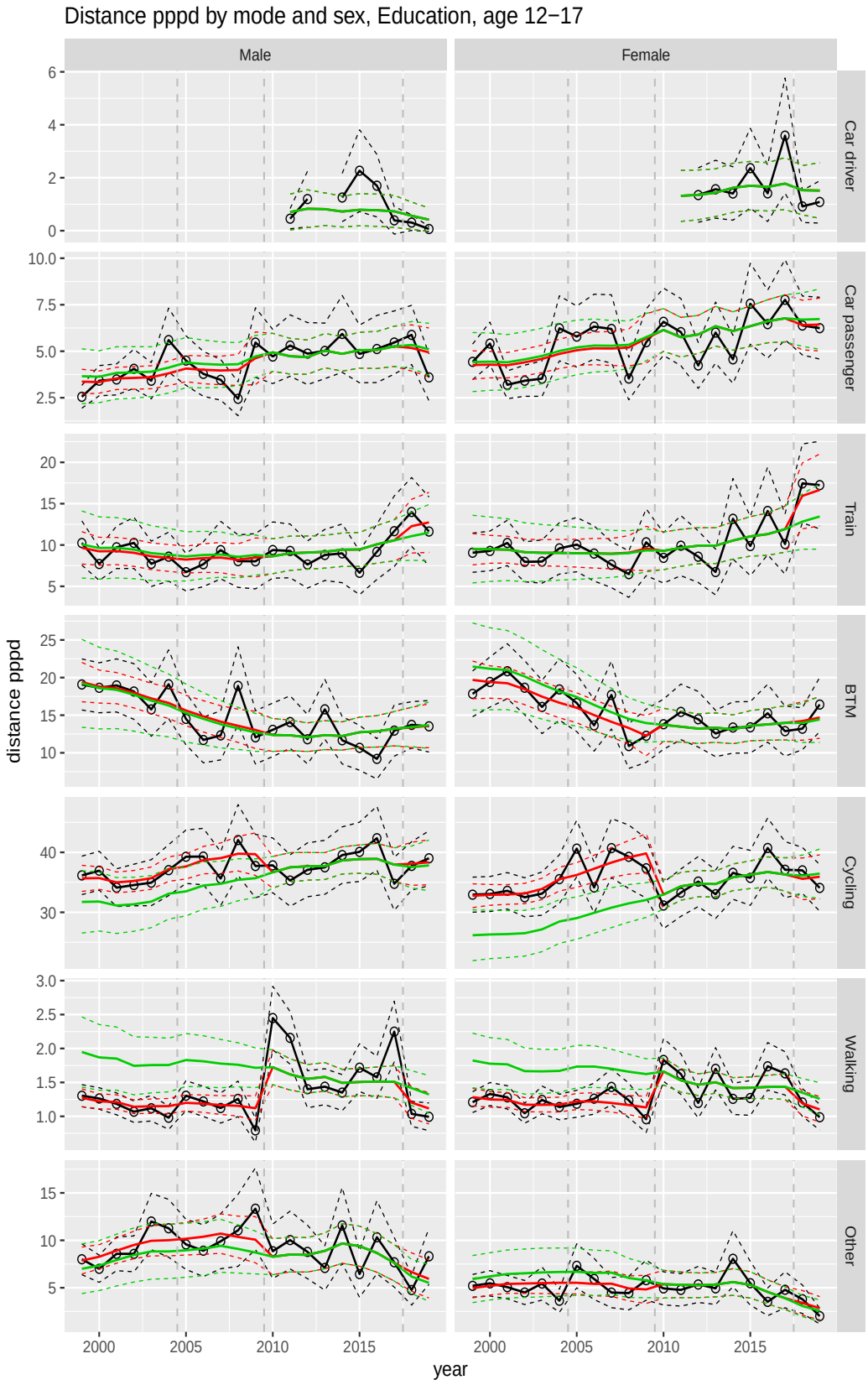

Figure A.188 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

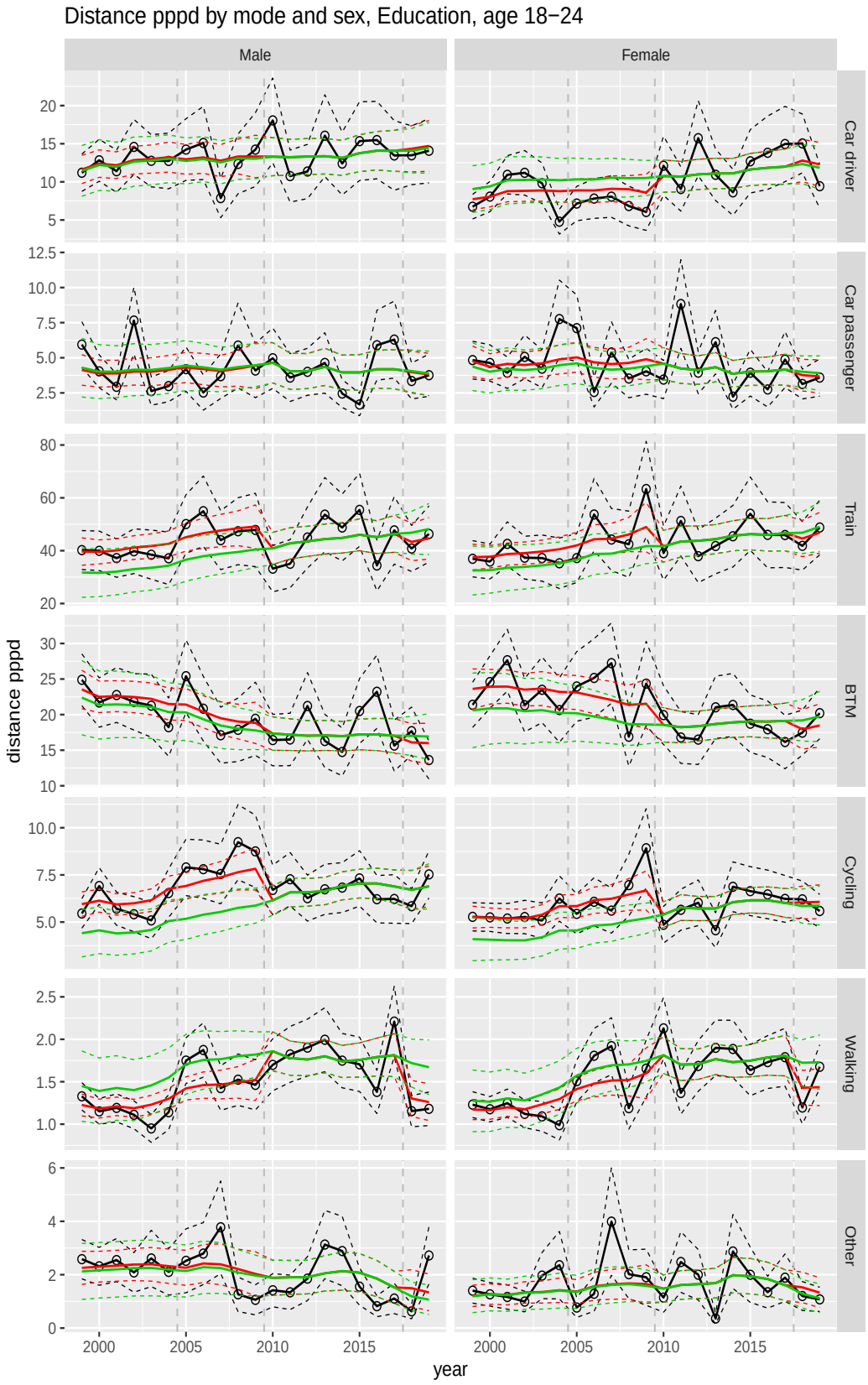

Figure A.189 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

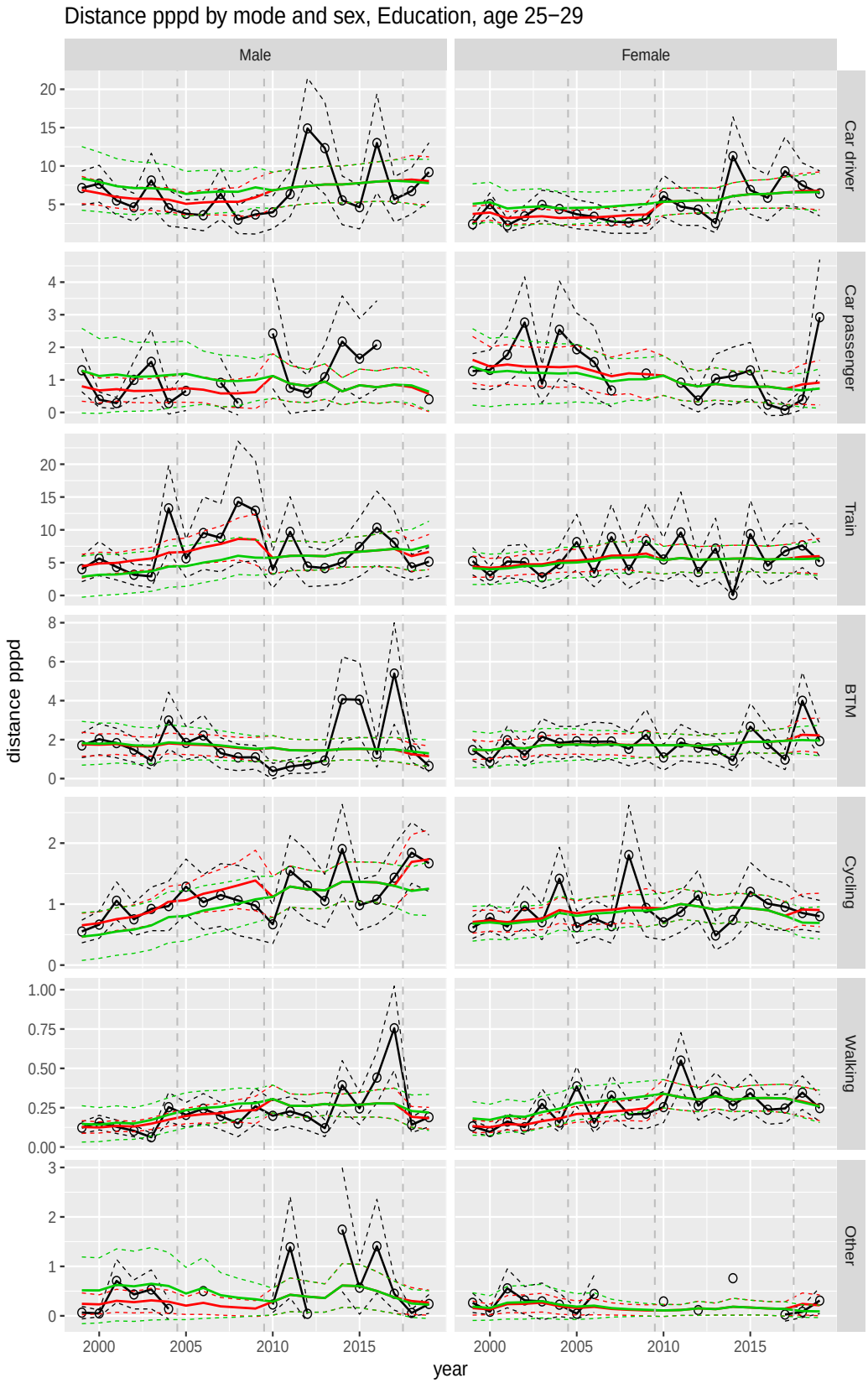

Figure A.190 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

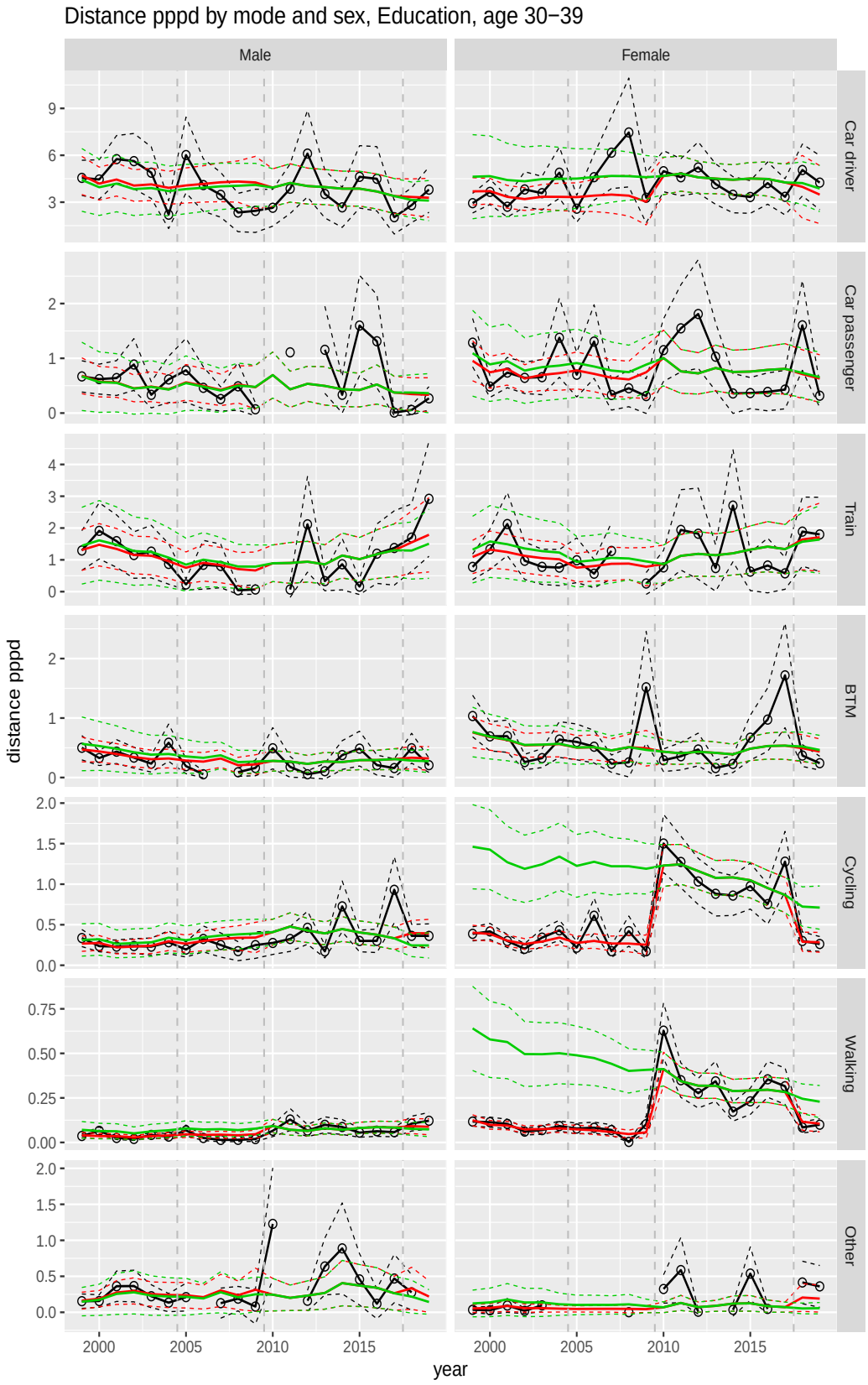

Figure A.191 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

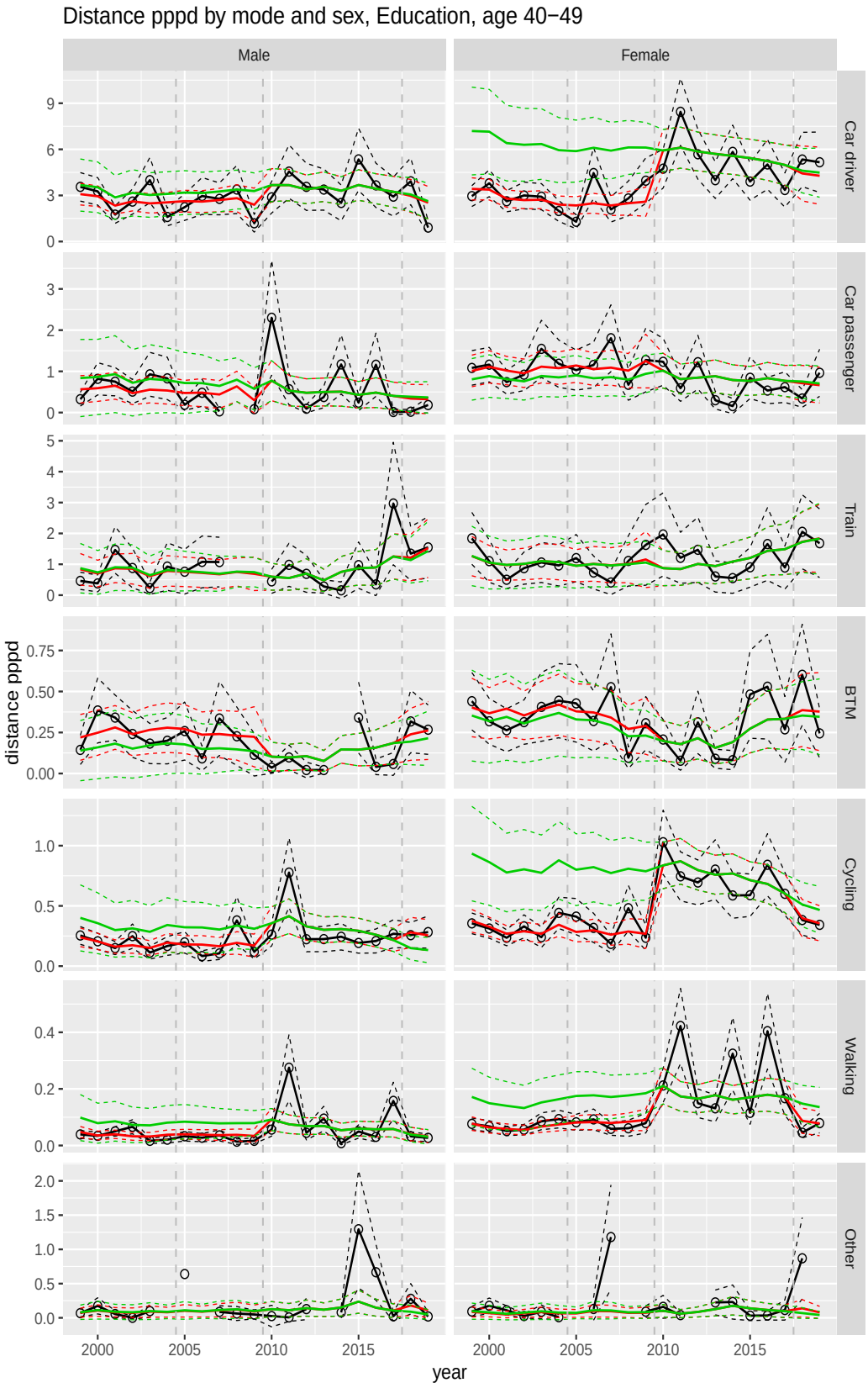

Figure A.192 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

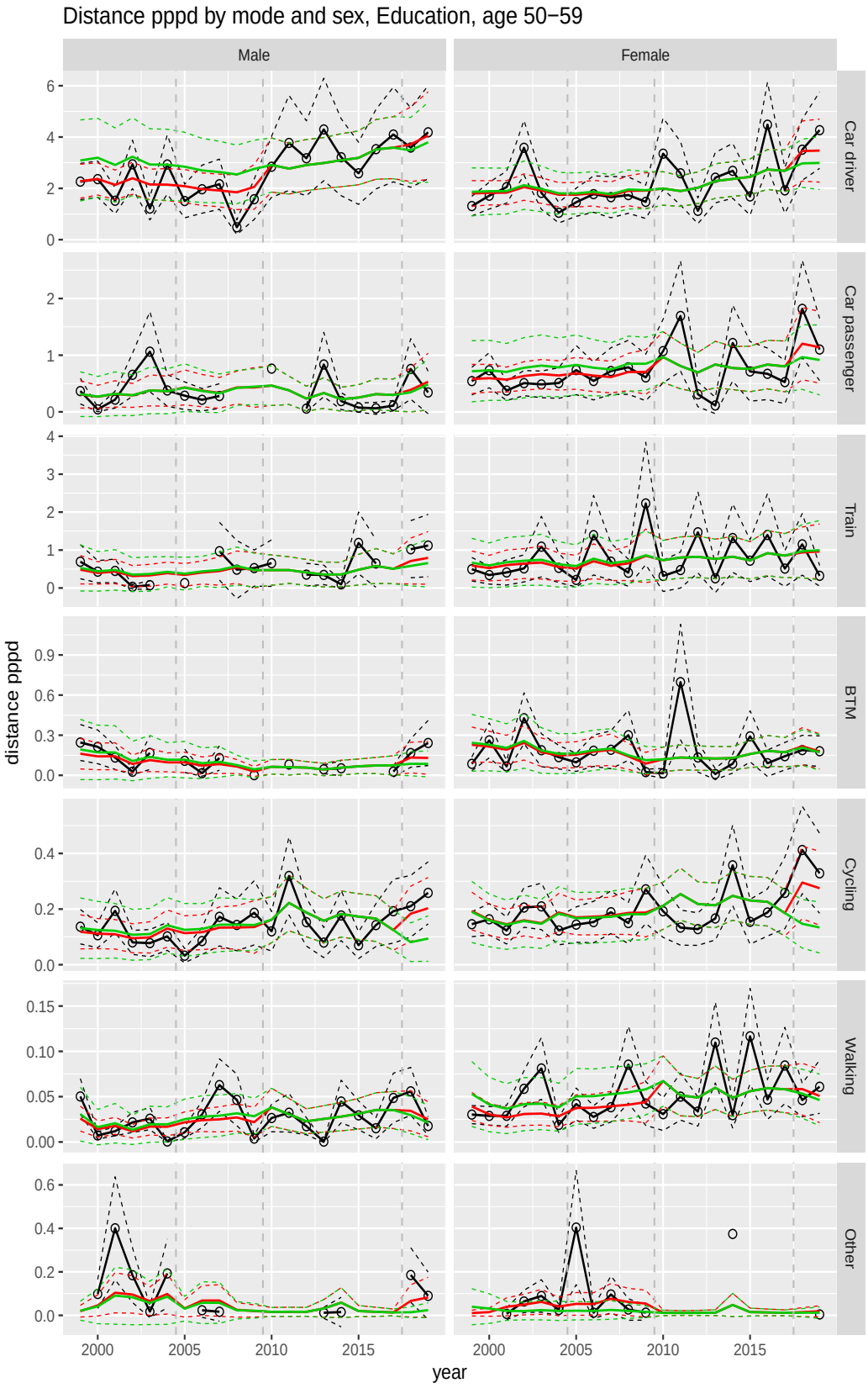

Figure A.193 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

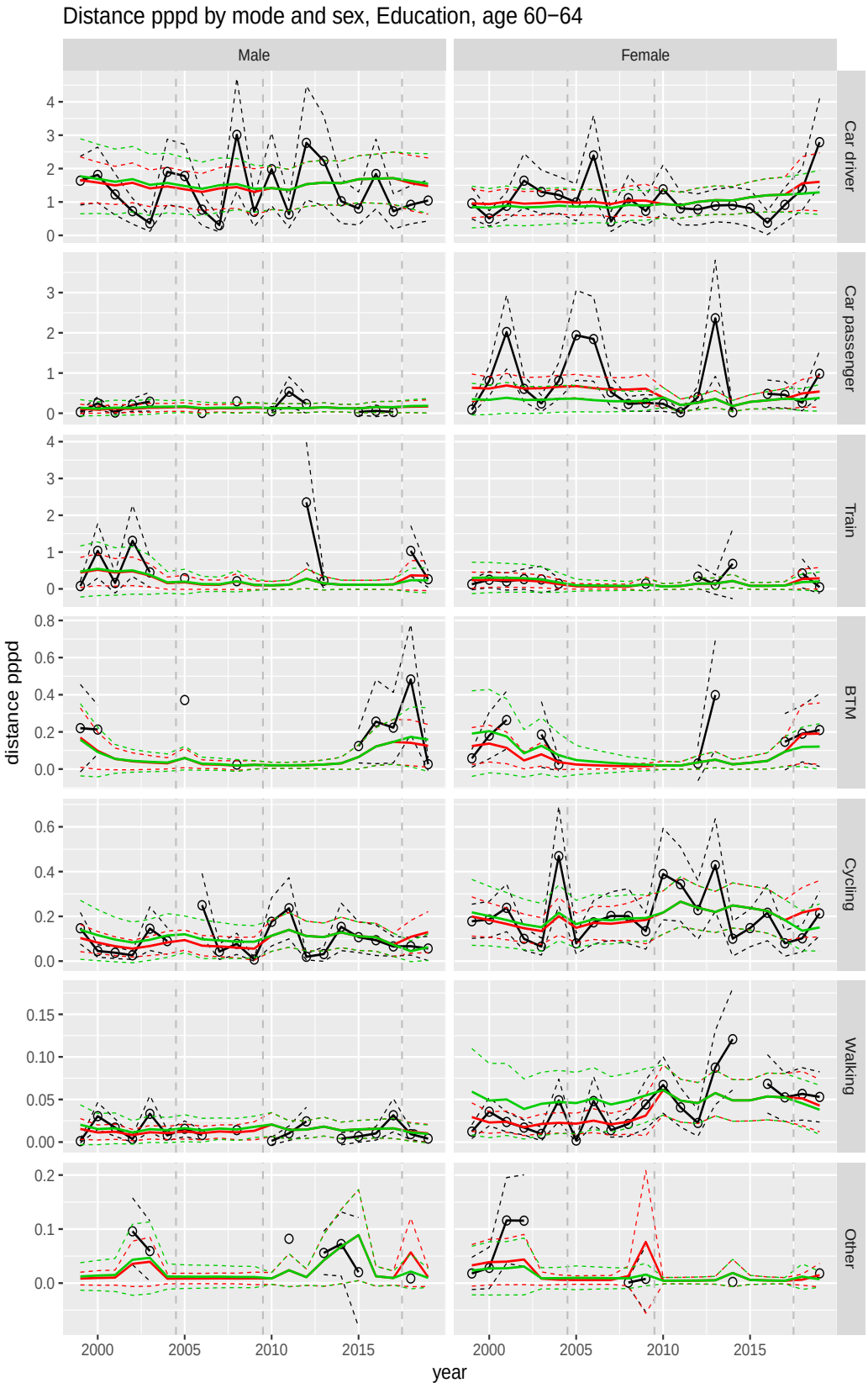

Figure A.194 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

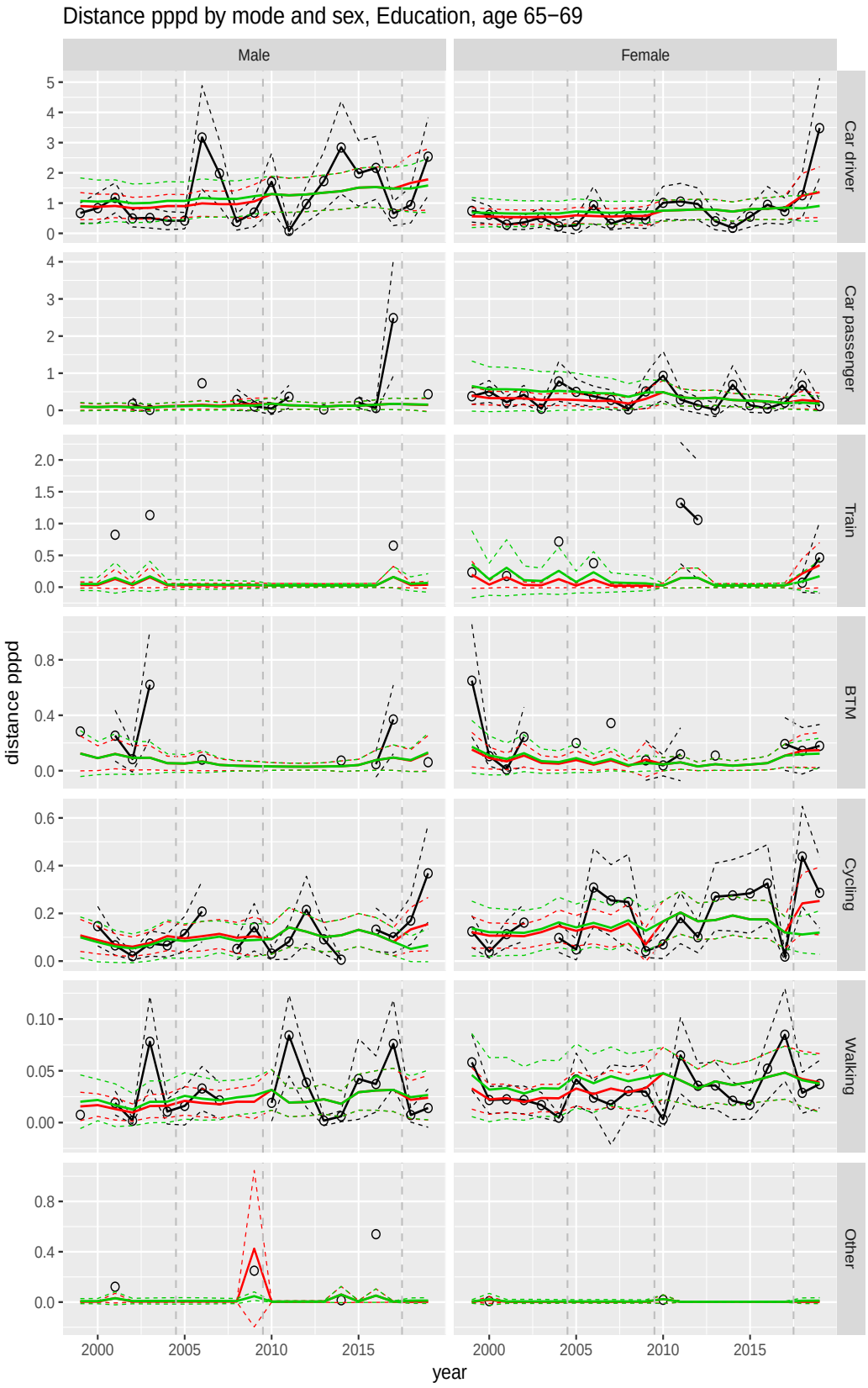

Figure A.195 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

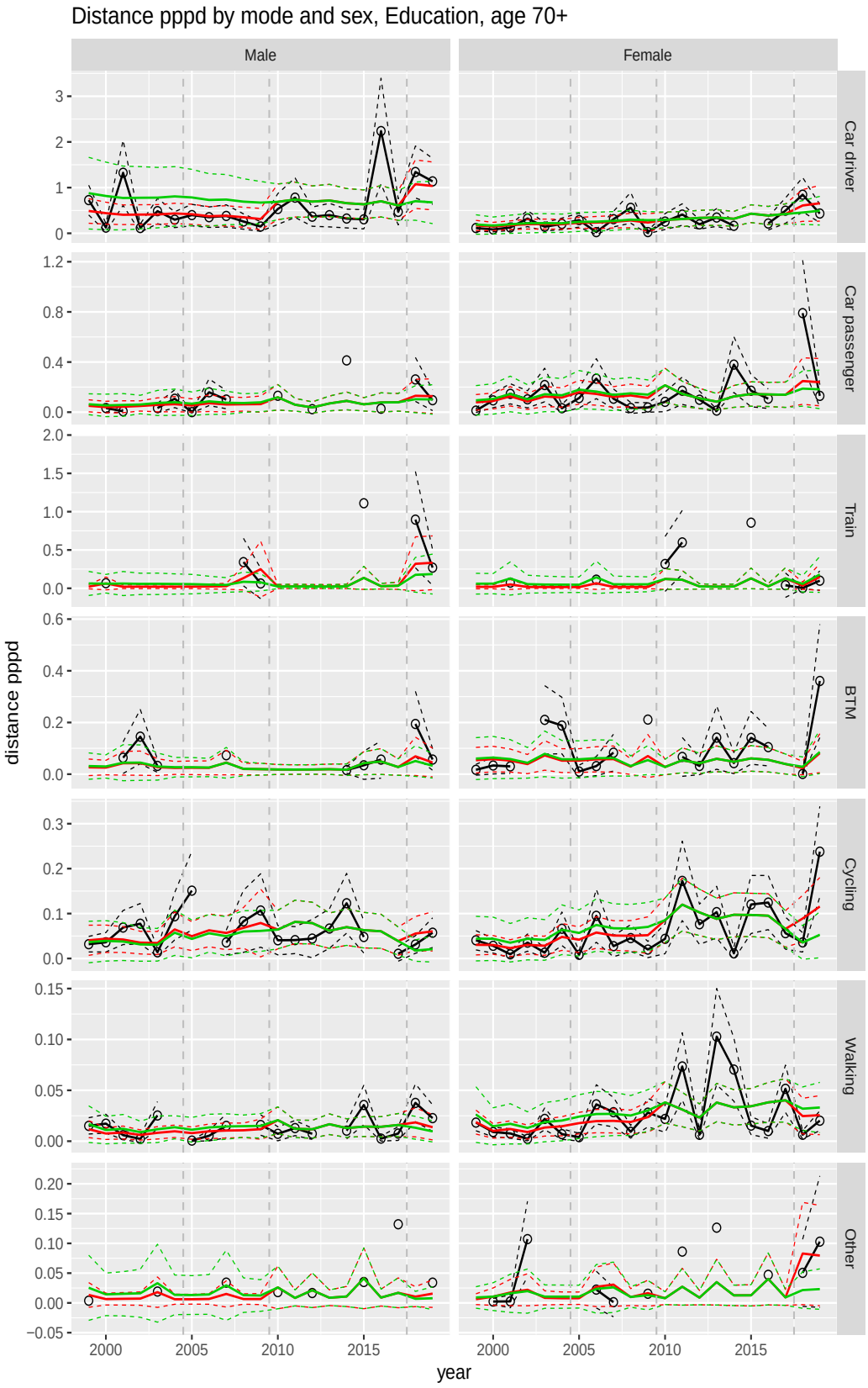

Figure A.196 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

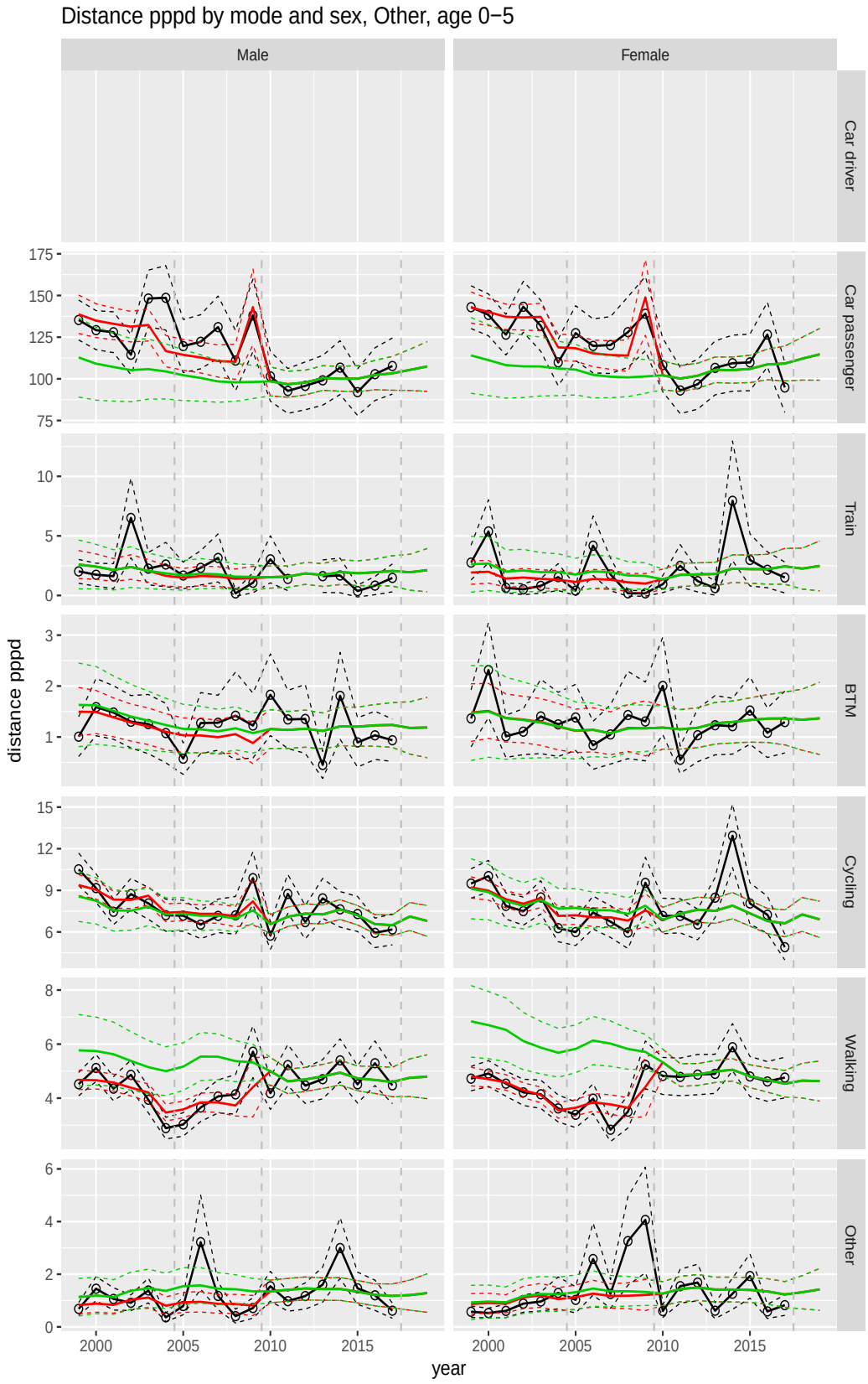

Figure A.197 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

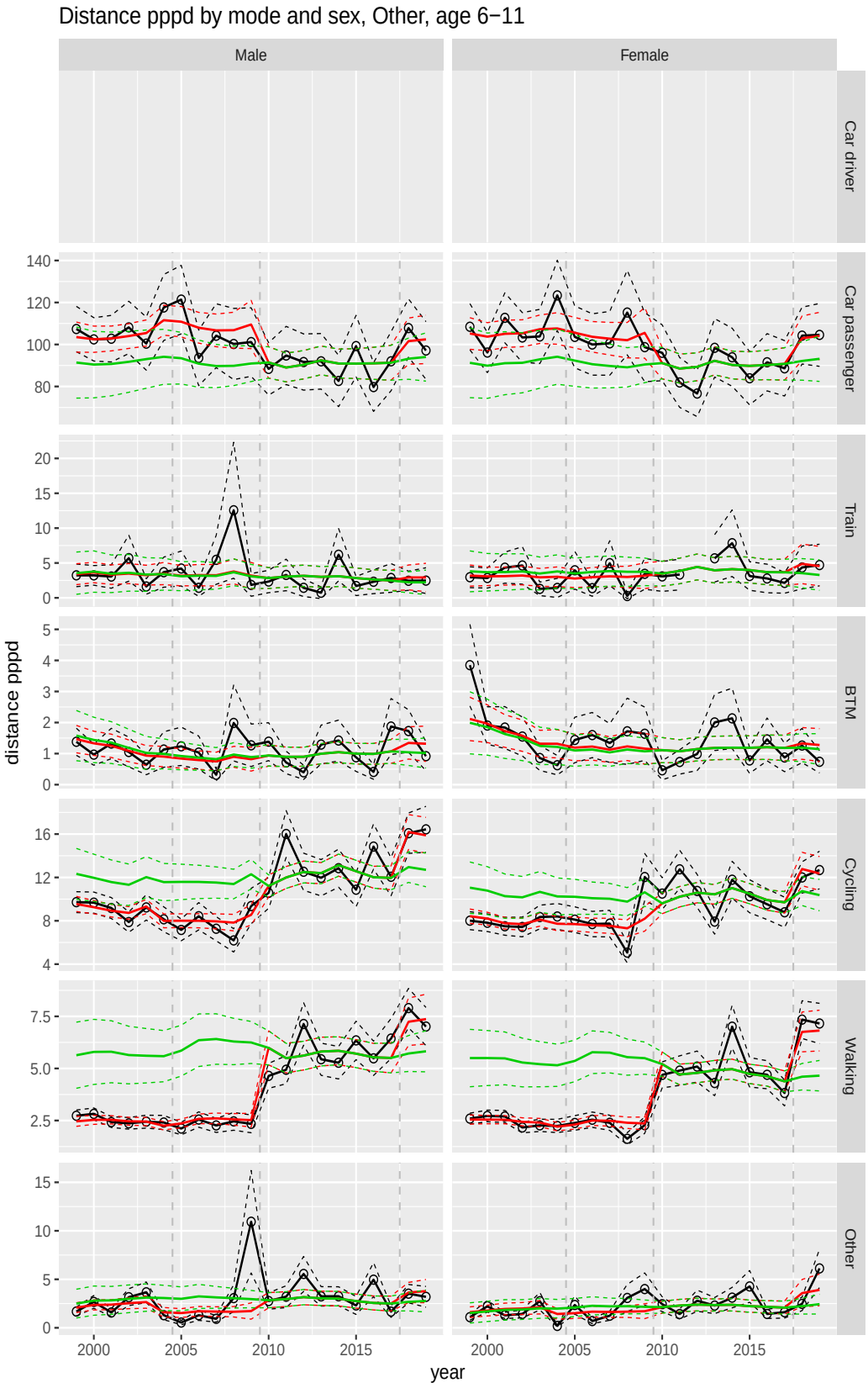

Figure A.198 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

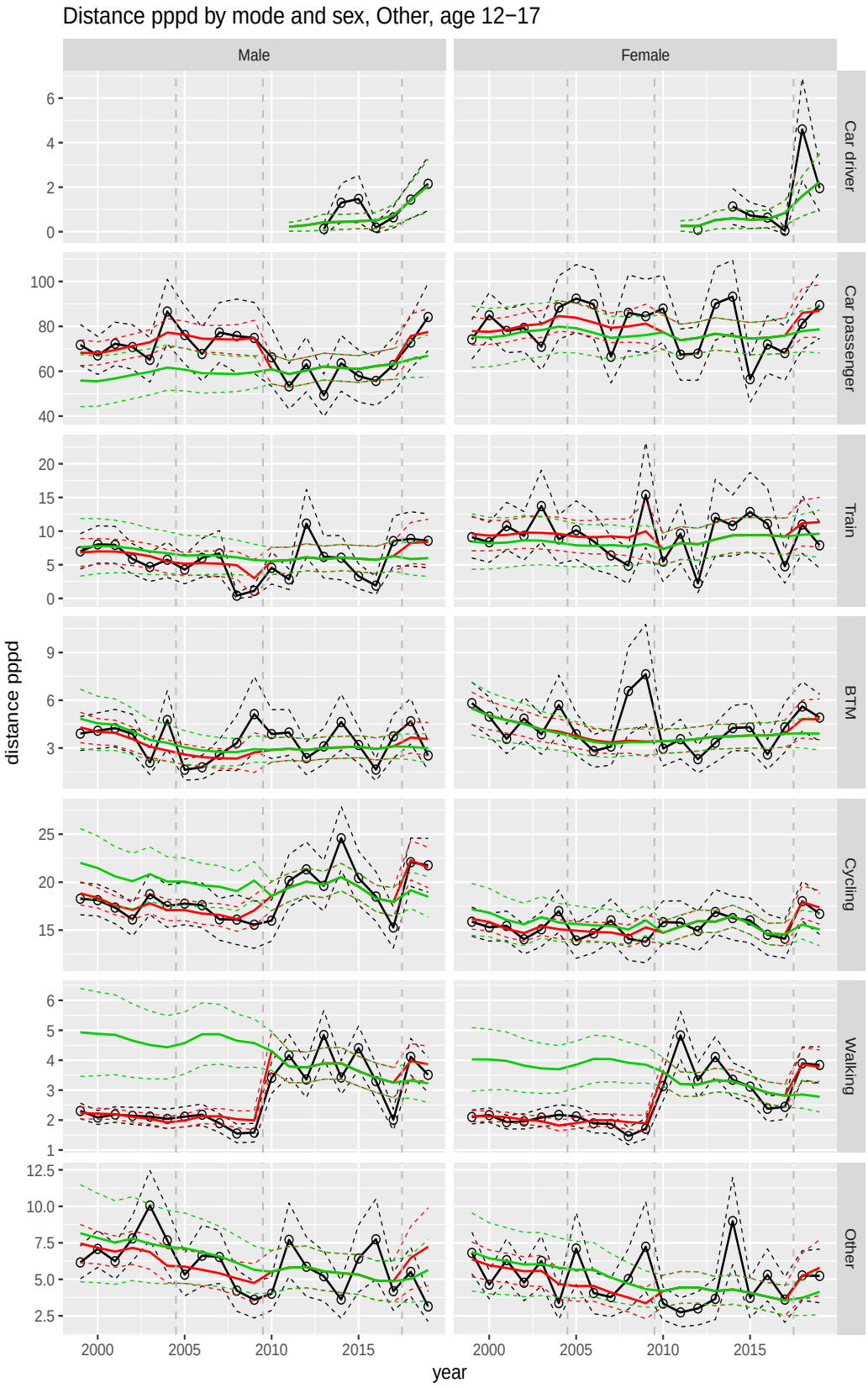

Figure A.199 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

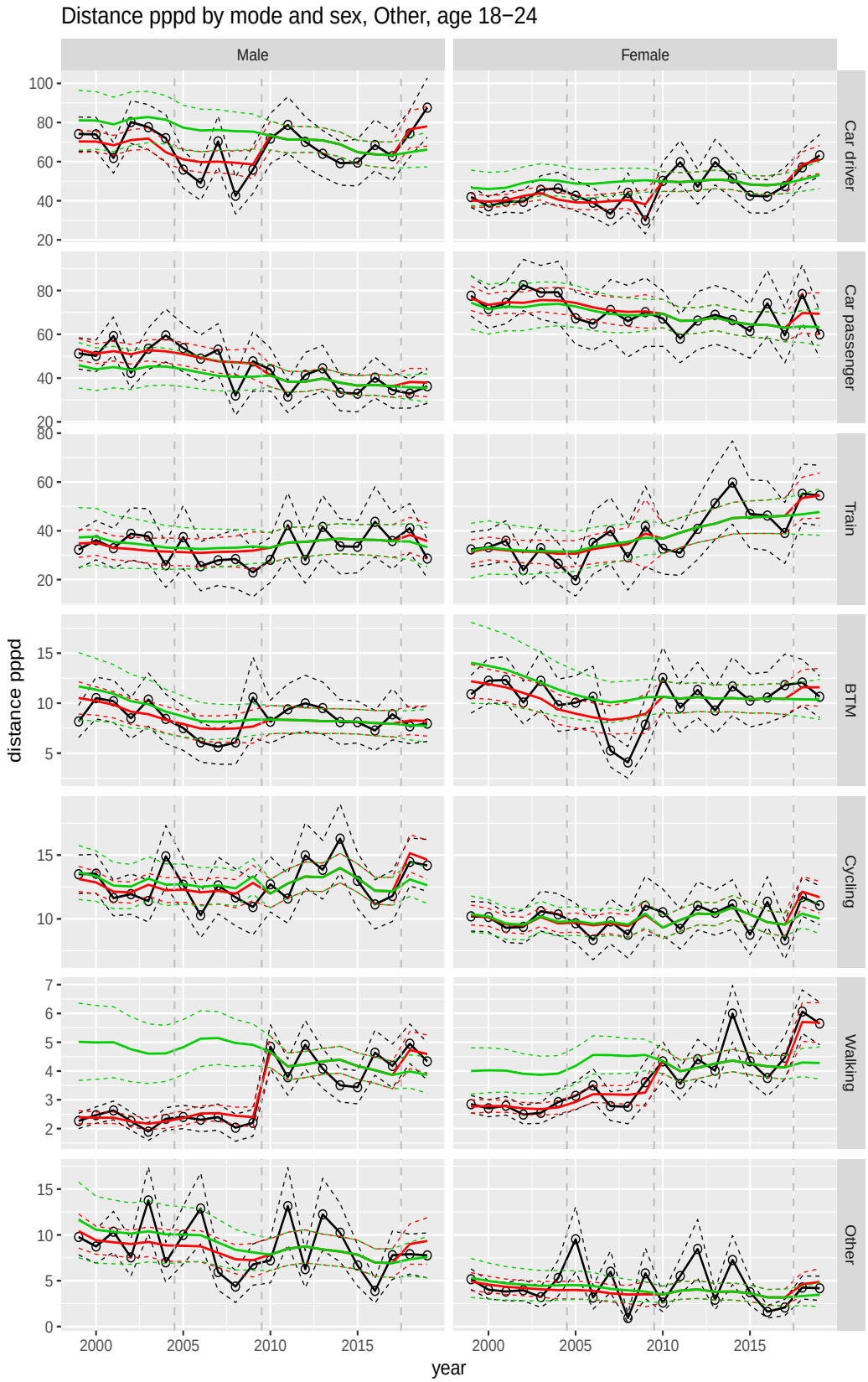

Figure A.200 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

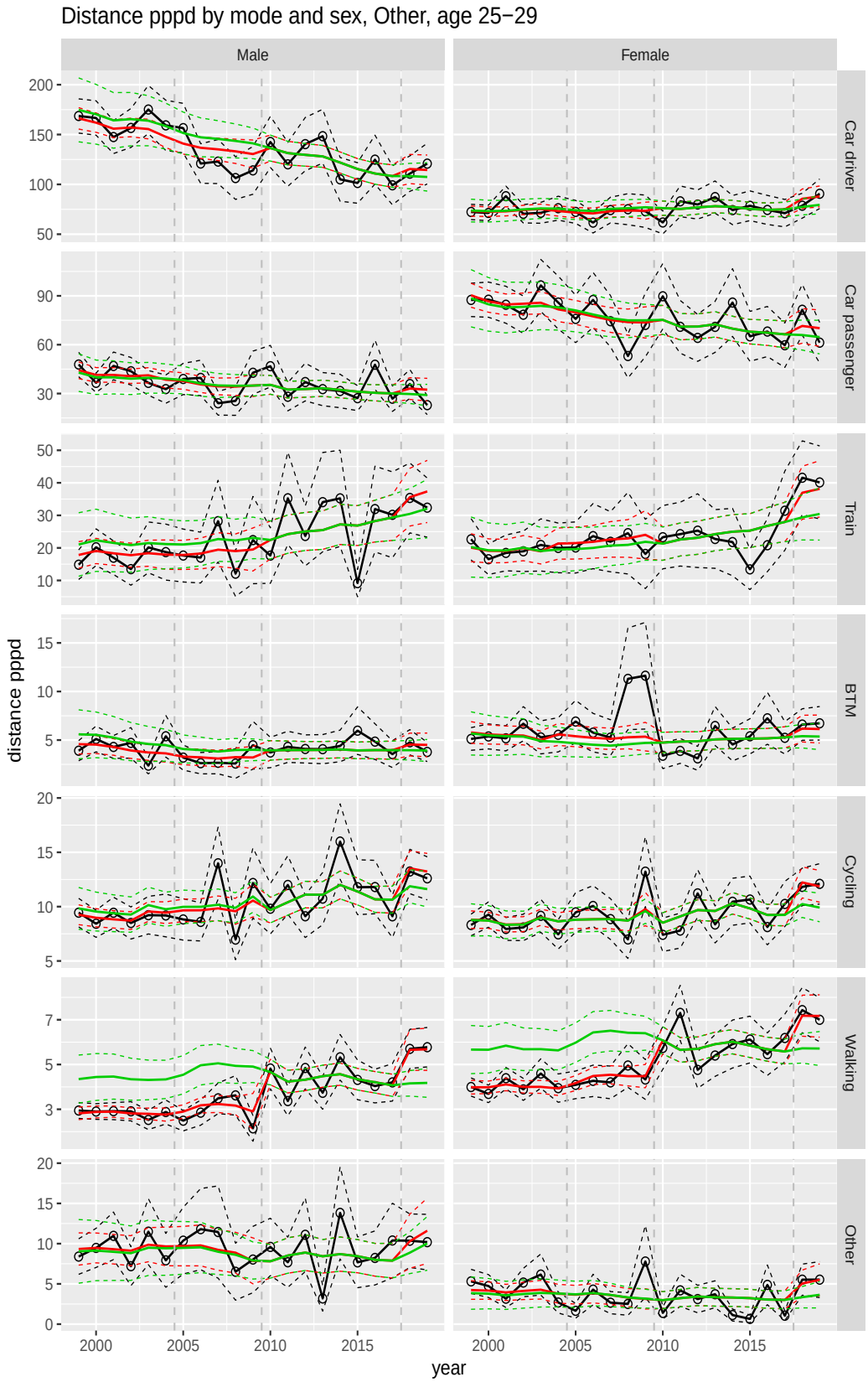

Figure A.201 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure A.202 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

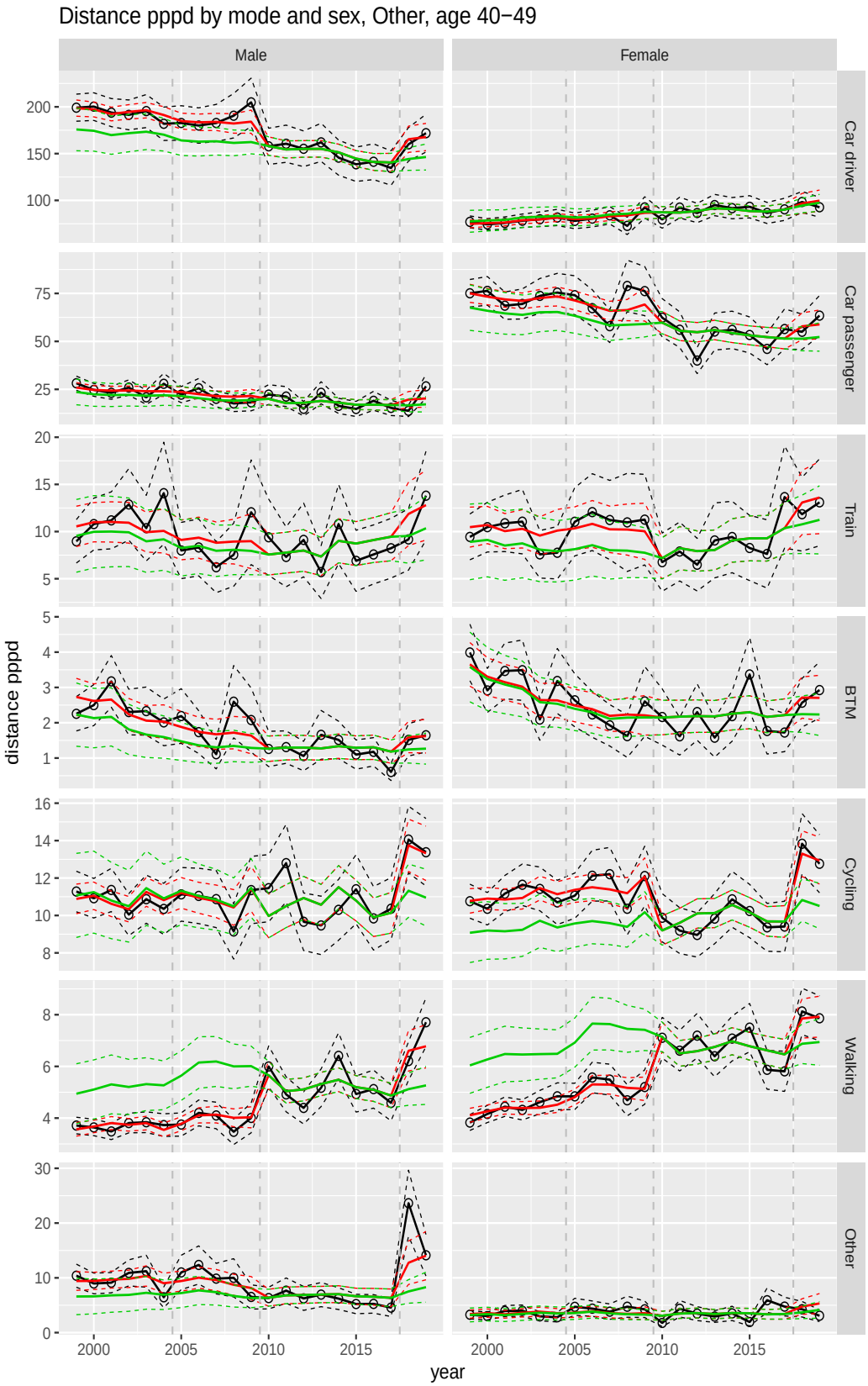

Figure A.203 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

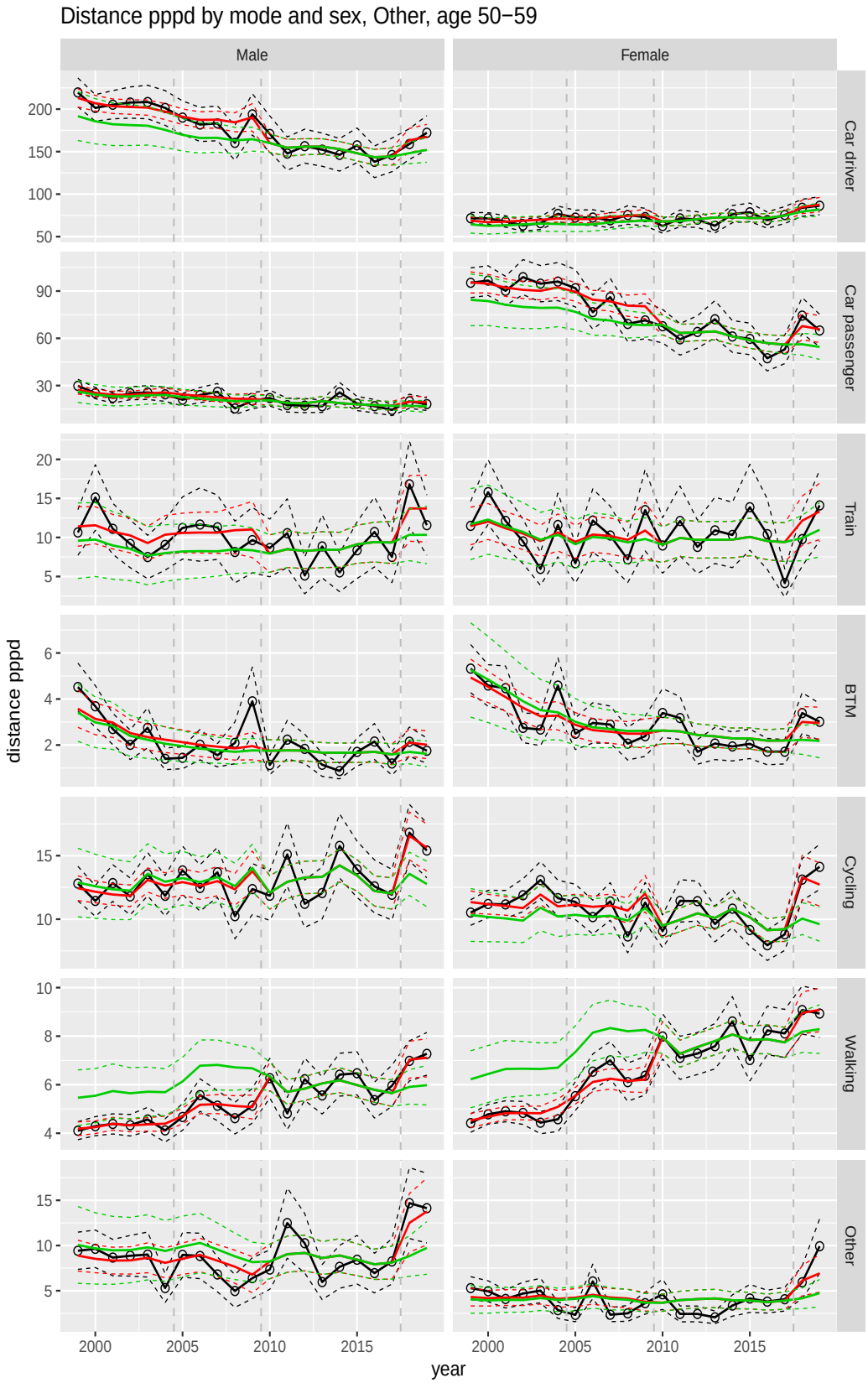

Figure A.204 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

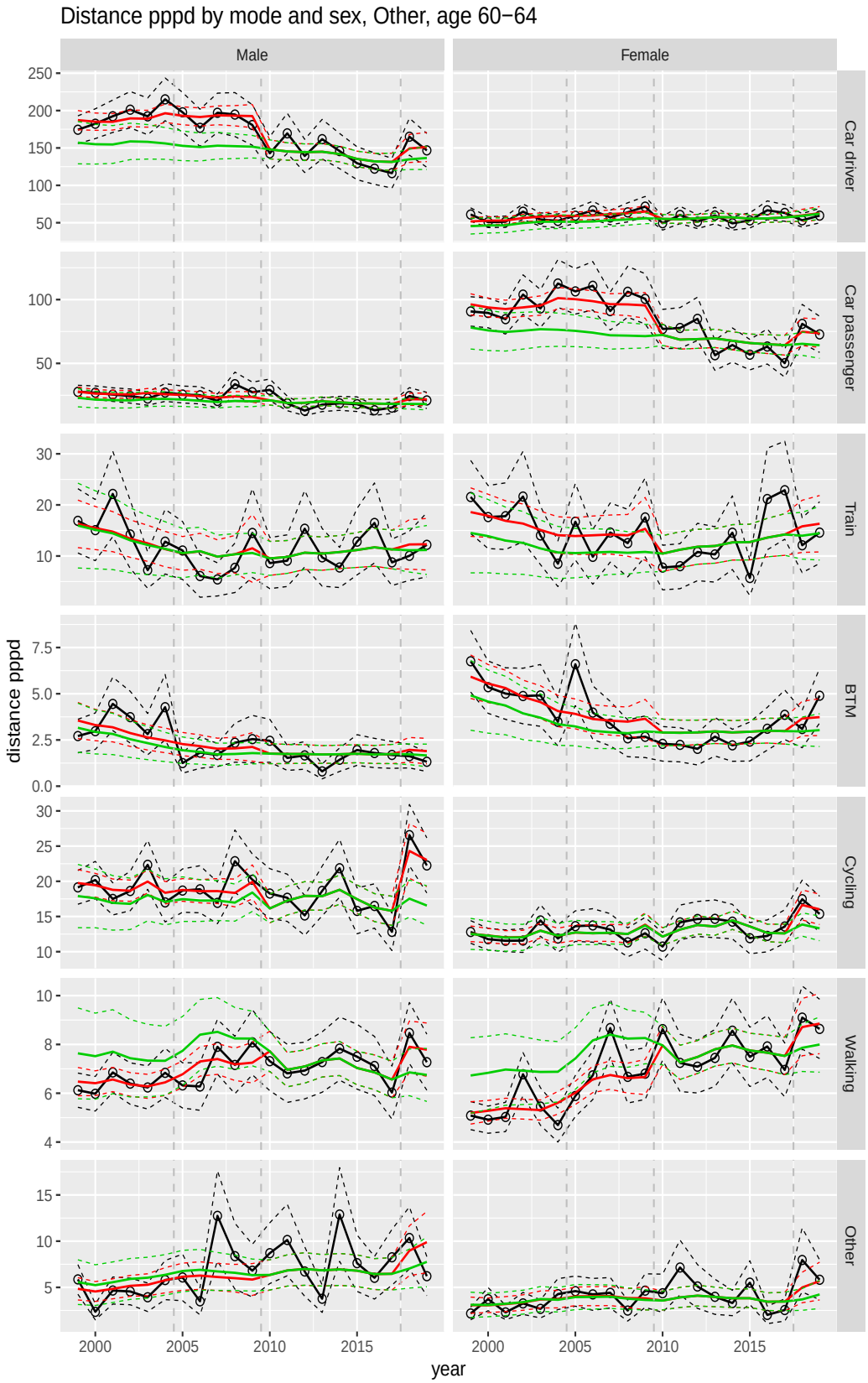

Figure A.205 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

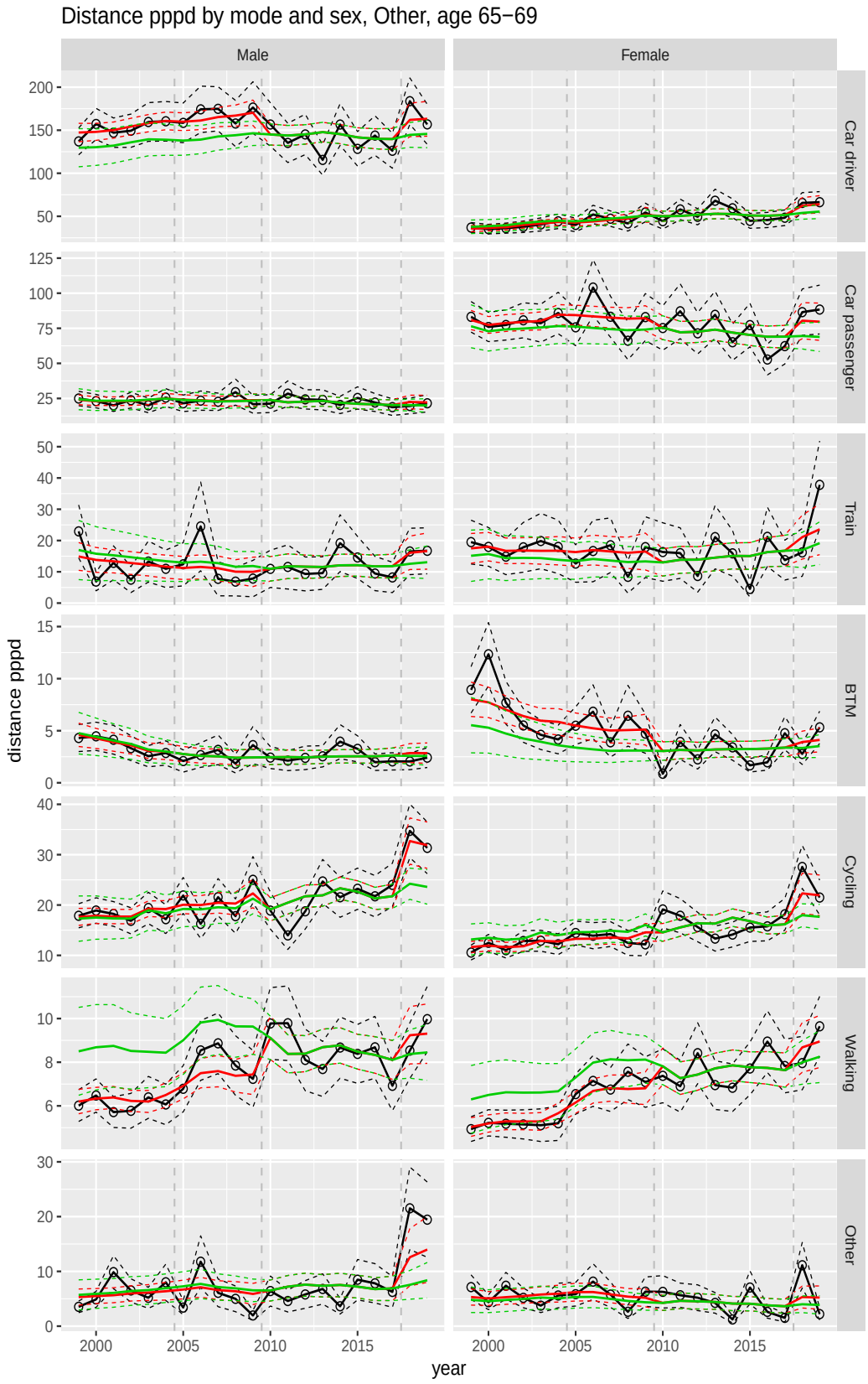

Figure A.206 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

Figure A.207 Direct estimates (black), model fit (red) and trend estimates (green) with approximate 95% intervals.

	k=1		k=2		k=3		k=4	
	OVIN	OVG	OVIN	OVG	OVIN	OVG		
t				$AR_k \times 100$				
2010	2.221	0.744	3.199	2.109	3.805	2.745	4.406	2.254
2011	0.107	2.225	0.736	2.767	1.119	2.534	1.284	2.705
2012	0.025	1.191	0.511	0.745	0.385	1.191	0.458	1.804
2013	0.418	0.388	0.357	0.007	0.771	0.266	1.082	0.166
2014	2.216	1.890	2.941	1.983	3.384	2.565		
2015	0.312	0.295	0.580	0.152				
2016	0.209	0.388						
t				$RR_k\%$				
2010	0.714	0.230	1.029	0.653	1.224	0.850	1.417	0.698
2011	0.036	0.708	0.246	0.880	0.374	0.806	0.429	0.860
2012	0.009	0.376	0.171	0.235	0.129	0.376	0.153	0.569
2013	0.139	0.121	0.119	0.002	0.256	0.083	0.360	0.052
2014	0.735	0.588	0.975	0.617	1.121	0.798		
2015	0.107	0.095	0.199	0.049				
2016	0.073	0.126						

Table A.1 Absolute Revision ($AR_k \times 100$) and Relative Revision ($RR_k\%$) of the signals at overall level for mean number of trip legs pppd for revision horizon $k = 1, 2, 3, 4$ when the signals are published at OViN and OVG level.

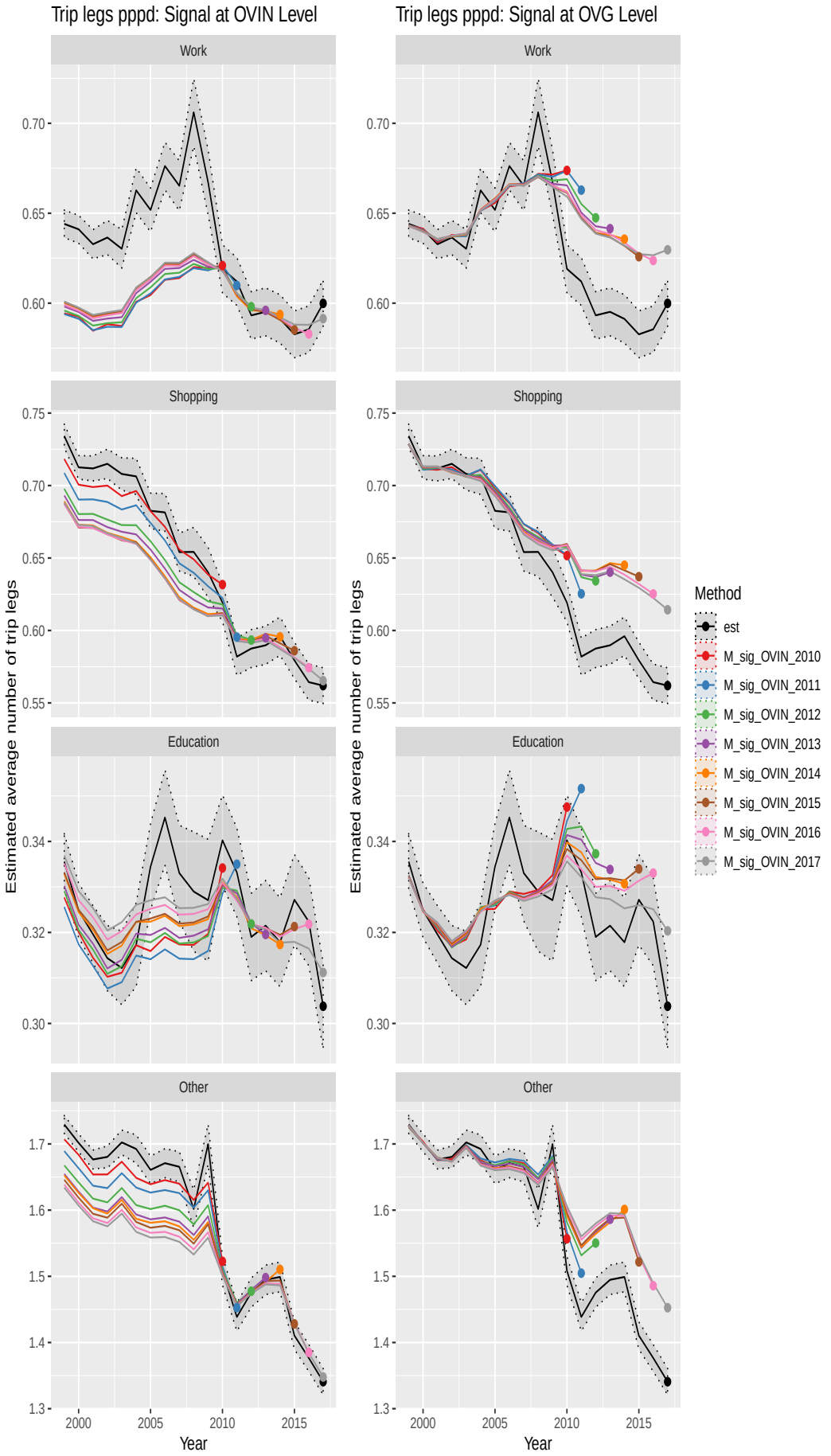

Figure A.208 Signal of mean number of trip legs for motive level measured at OVIN and OVG levels

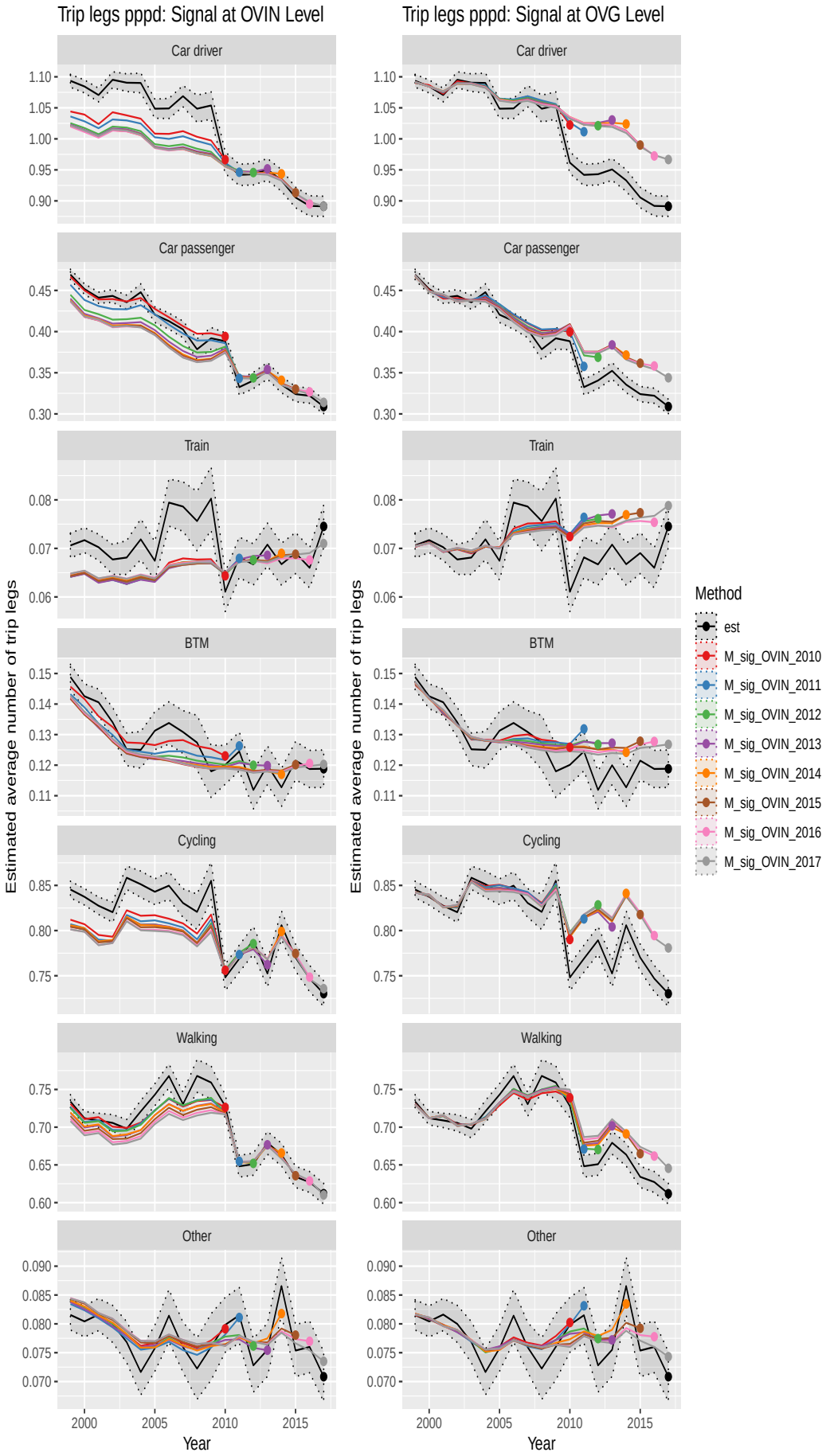

Figure A.209 Signal of mean number of trip legs for mode level measured at OVIN and OVG levels

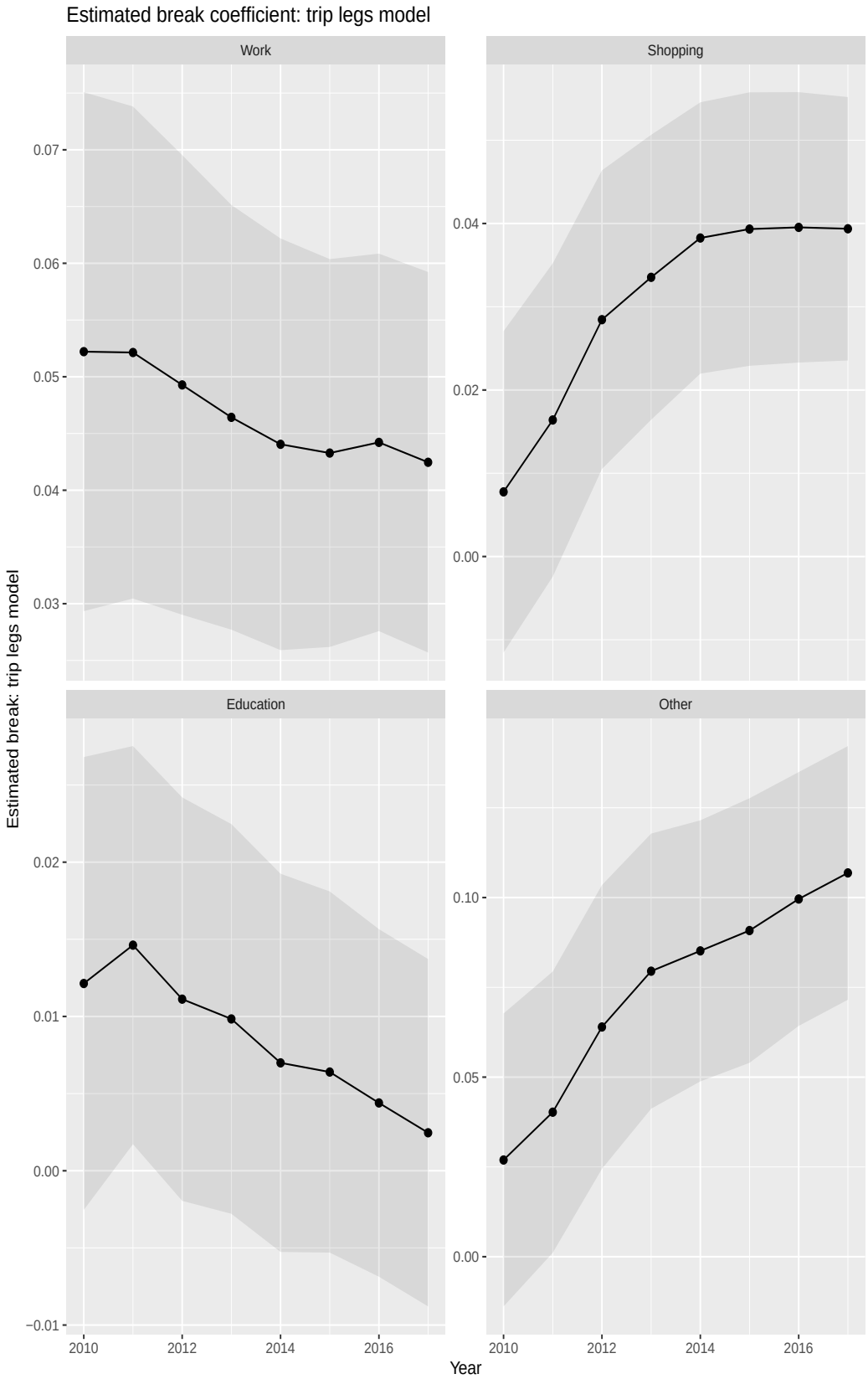

Figure A.210 In real time estimates from 2010 to 2017 for the discontinuity for number of trip legs at the motive level with approximate 95% confidence interval

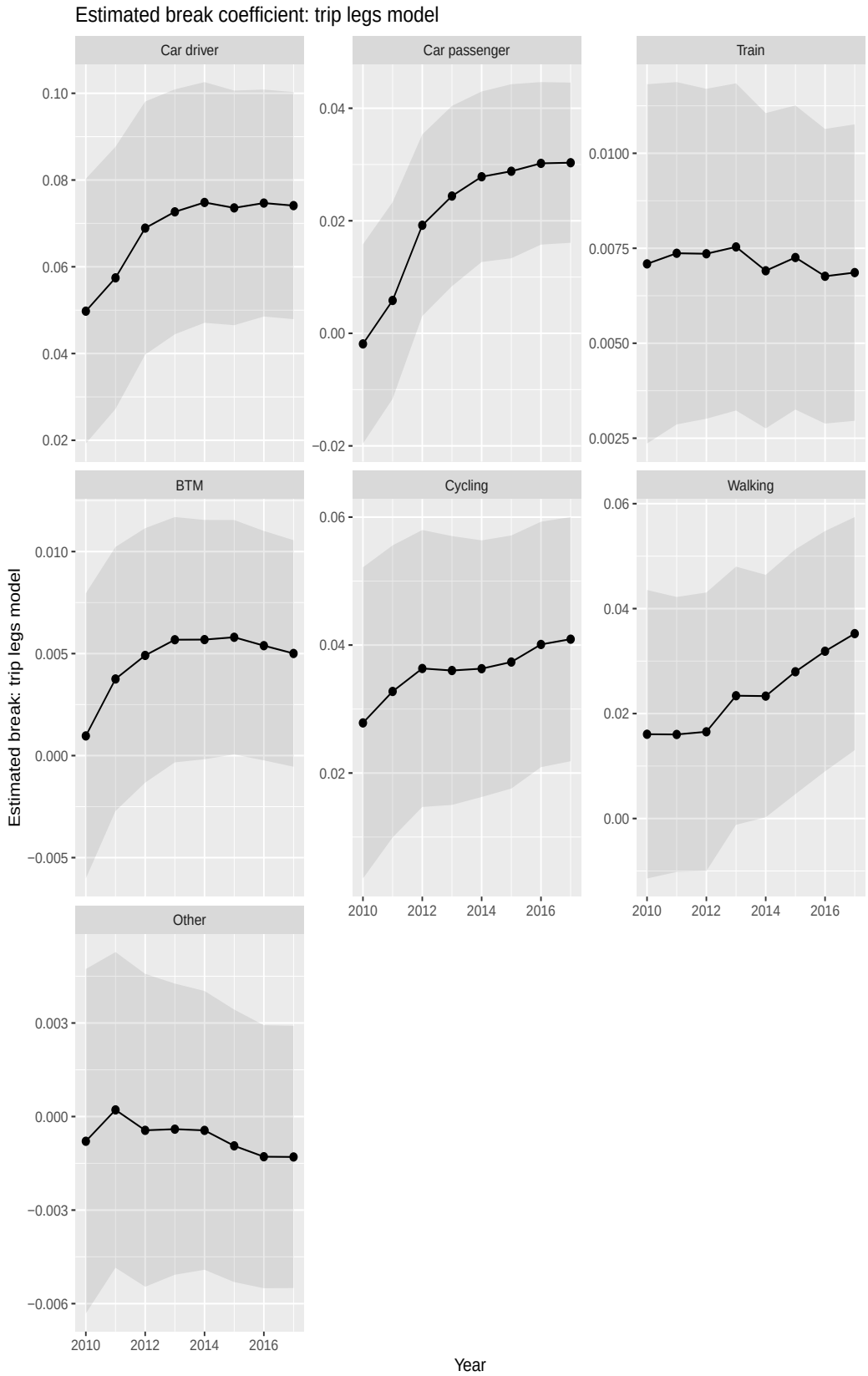

Figure A.211 In real time estimates from 2010 to 2017 for the discontinuity for number of trip legs at the mode level with approximate 95% confidence interval

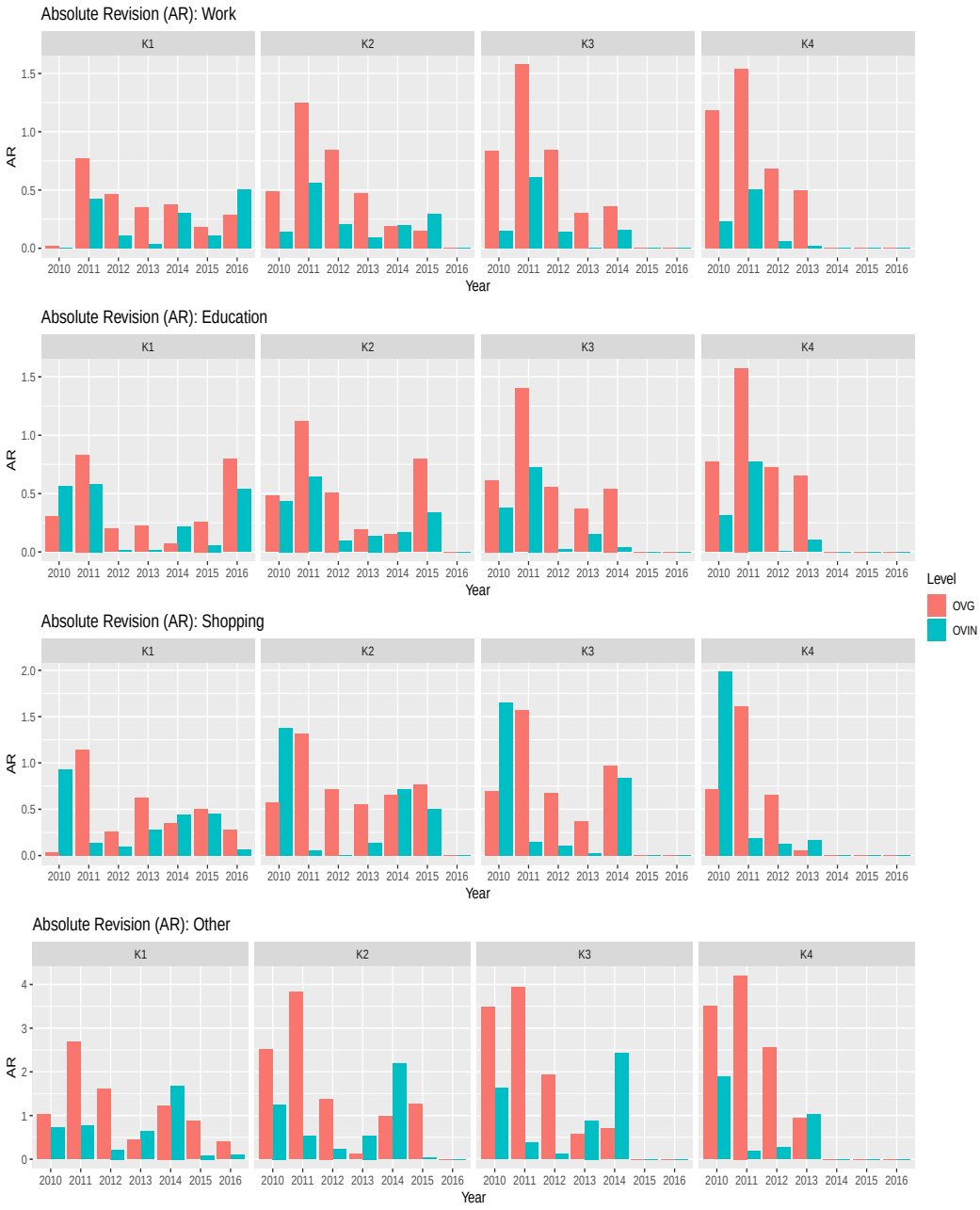

Figure A.212 Absolute Revision ($AR_k \times 100$) of the signals at motive levels for mean number of trip legs pppd when the signals are published at OViN and OVG level for revision horizon $k = 1, 2, 3, 4$.

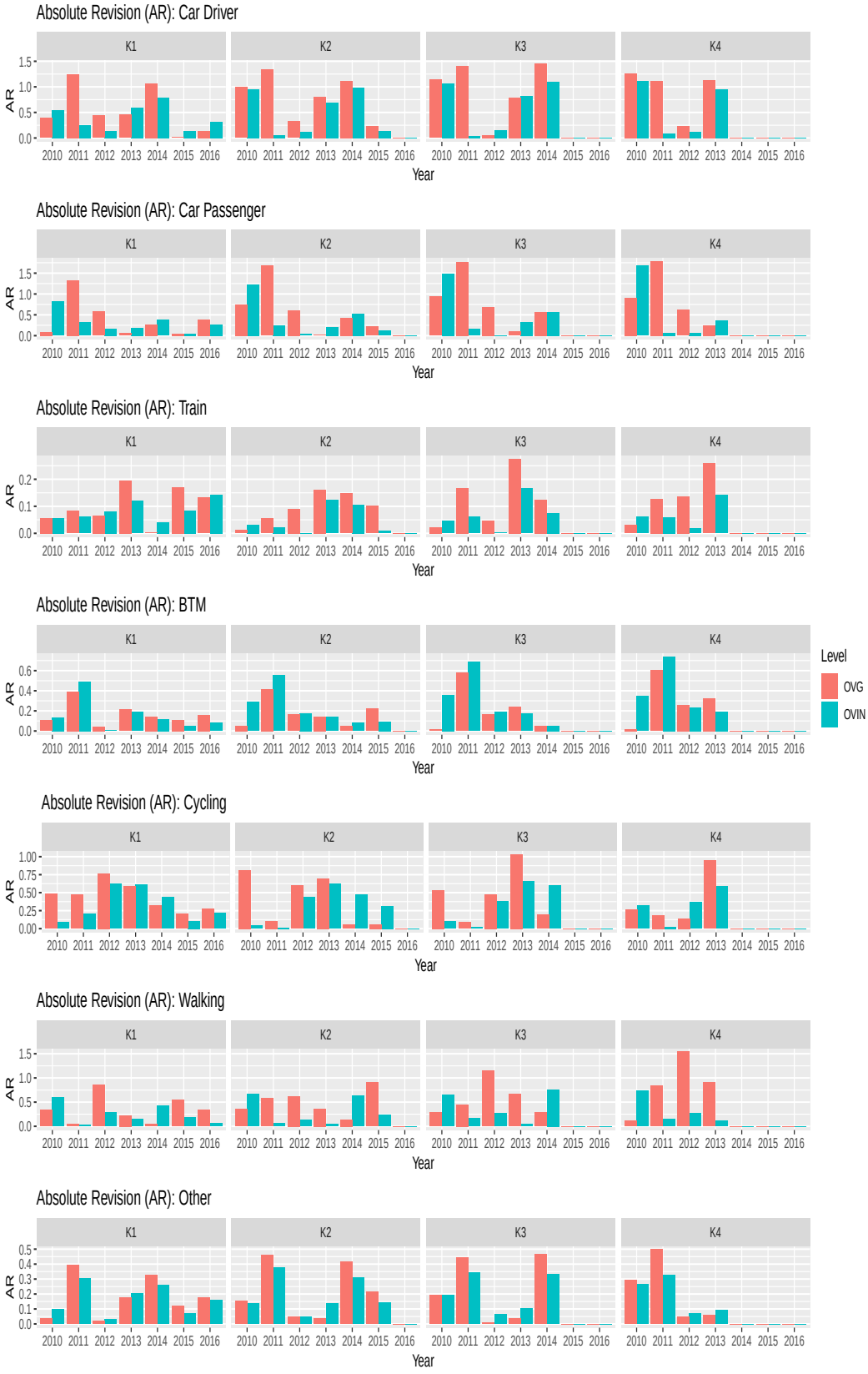

Figure A.213 $AR_k \times 100$**) of the signals at mode levels for mean number of trip legs pppd when the signals are published at OViN and OVG level for revision horizon** $k = 1, 2, 3, 4$**.**

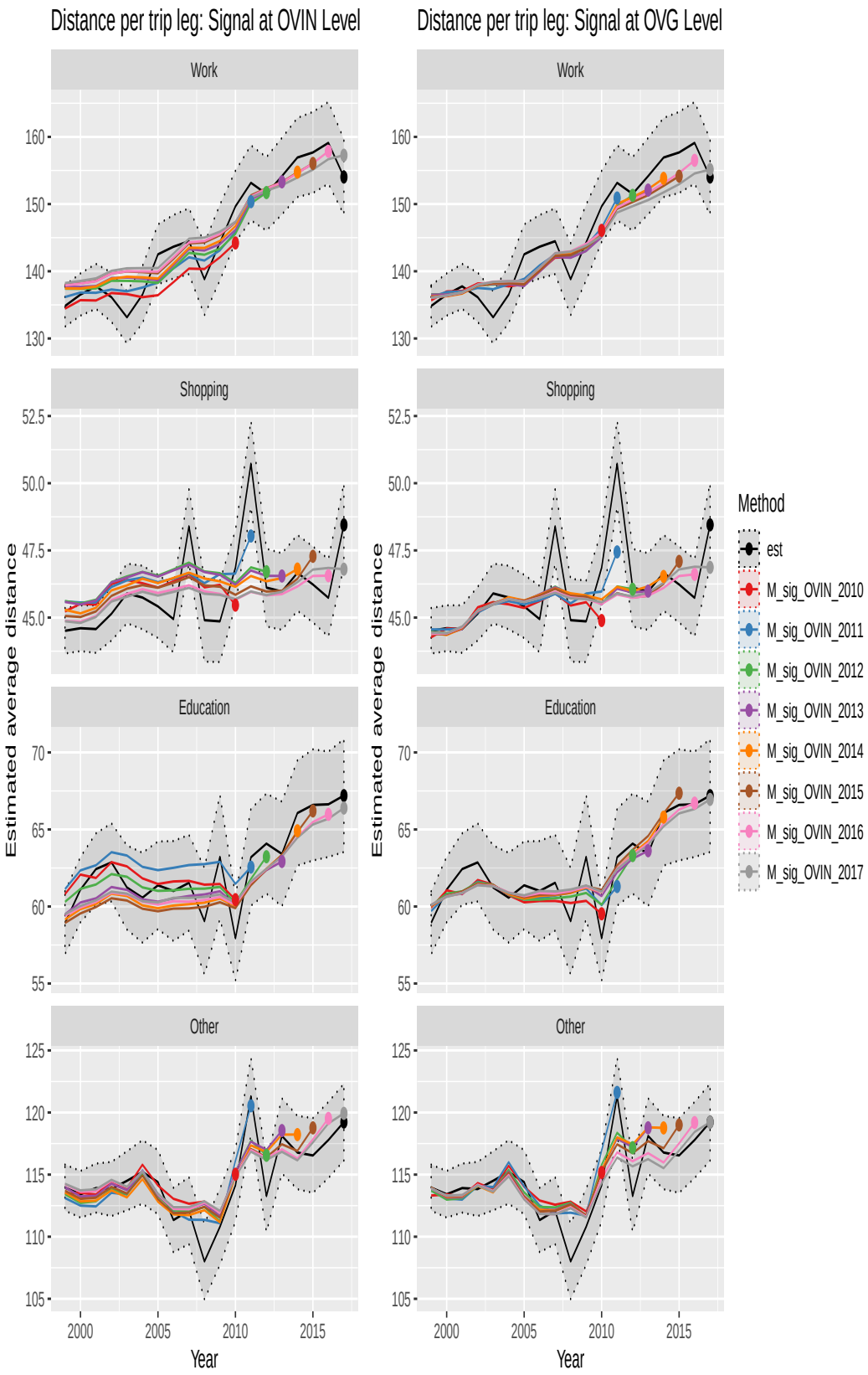

Figure A.214 Signal of average distance of trip legs for motive level measured at OVIN and OVG levels

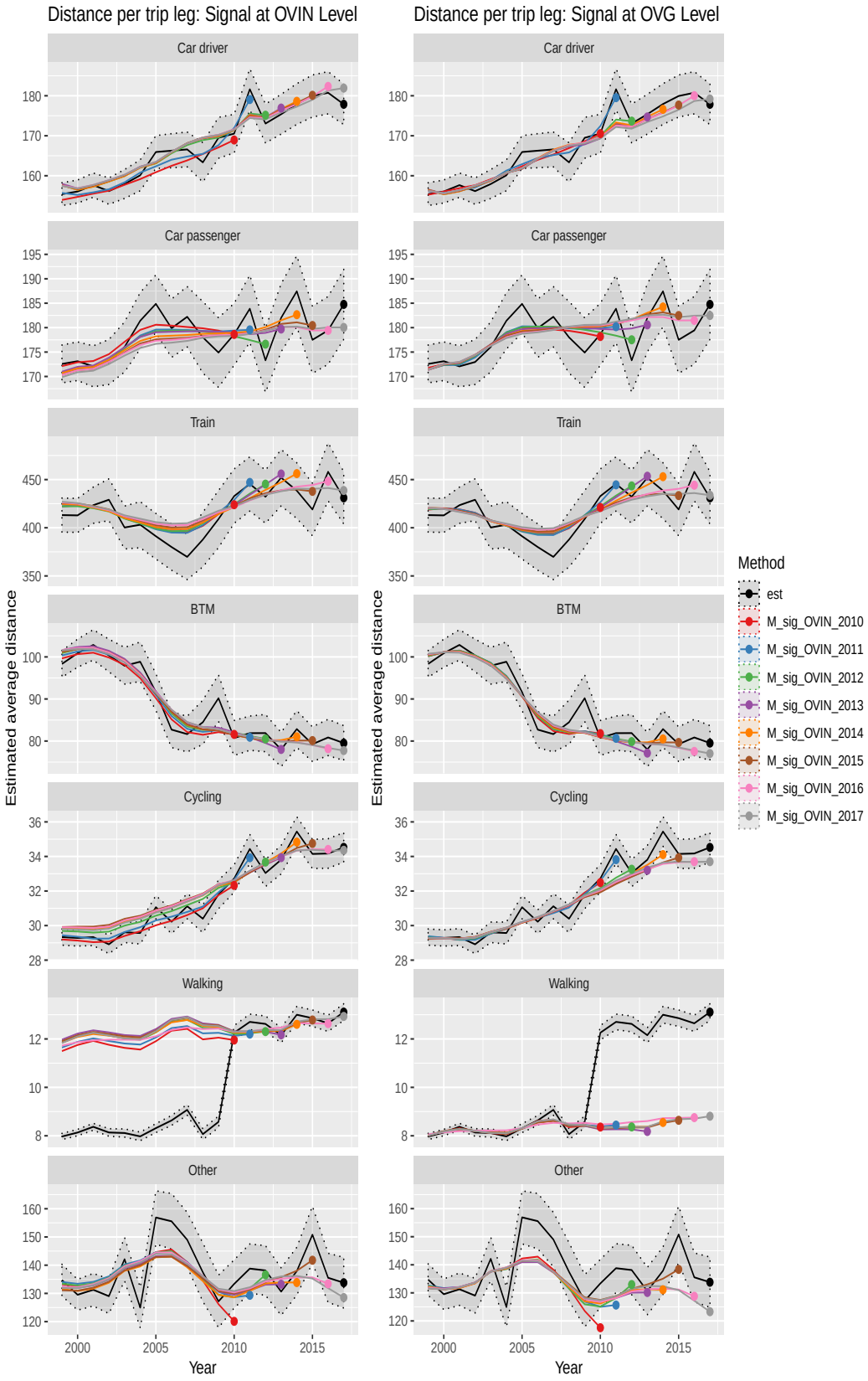

Figure A.215 Signal of average distance of trip legs for mode level measured at OVIN and OVG levels

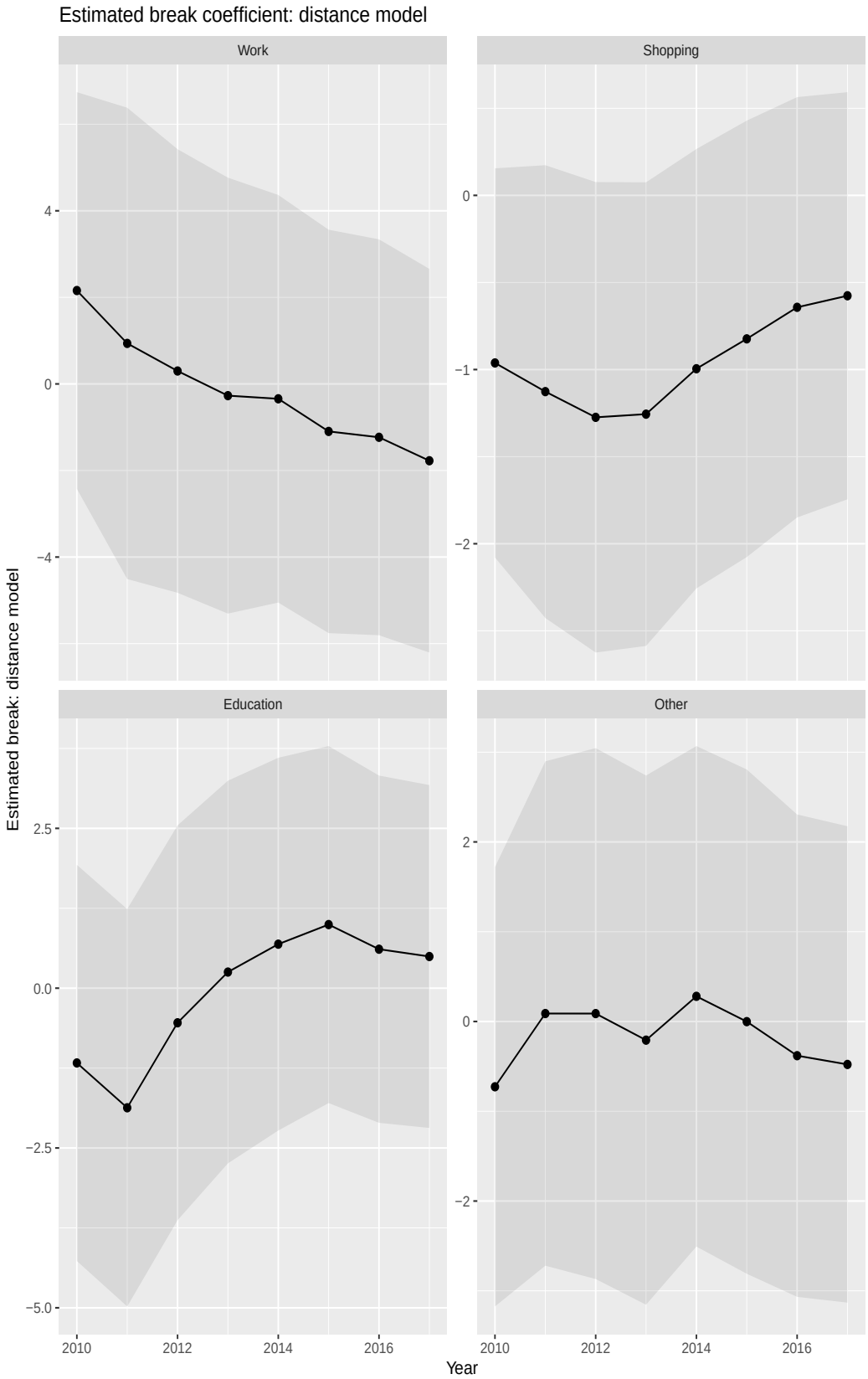

Figure A.216 In real time estimates from 2010 to 2017 for the discontinuity for number of trip legs at the motive level with approximate 95% confidence interval

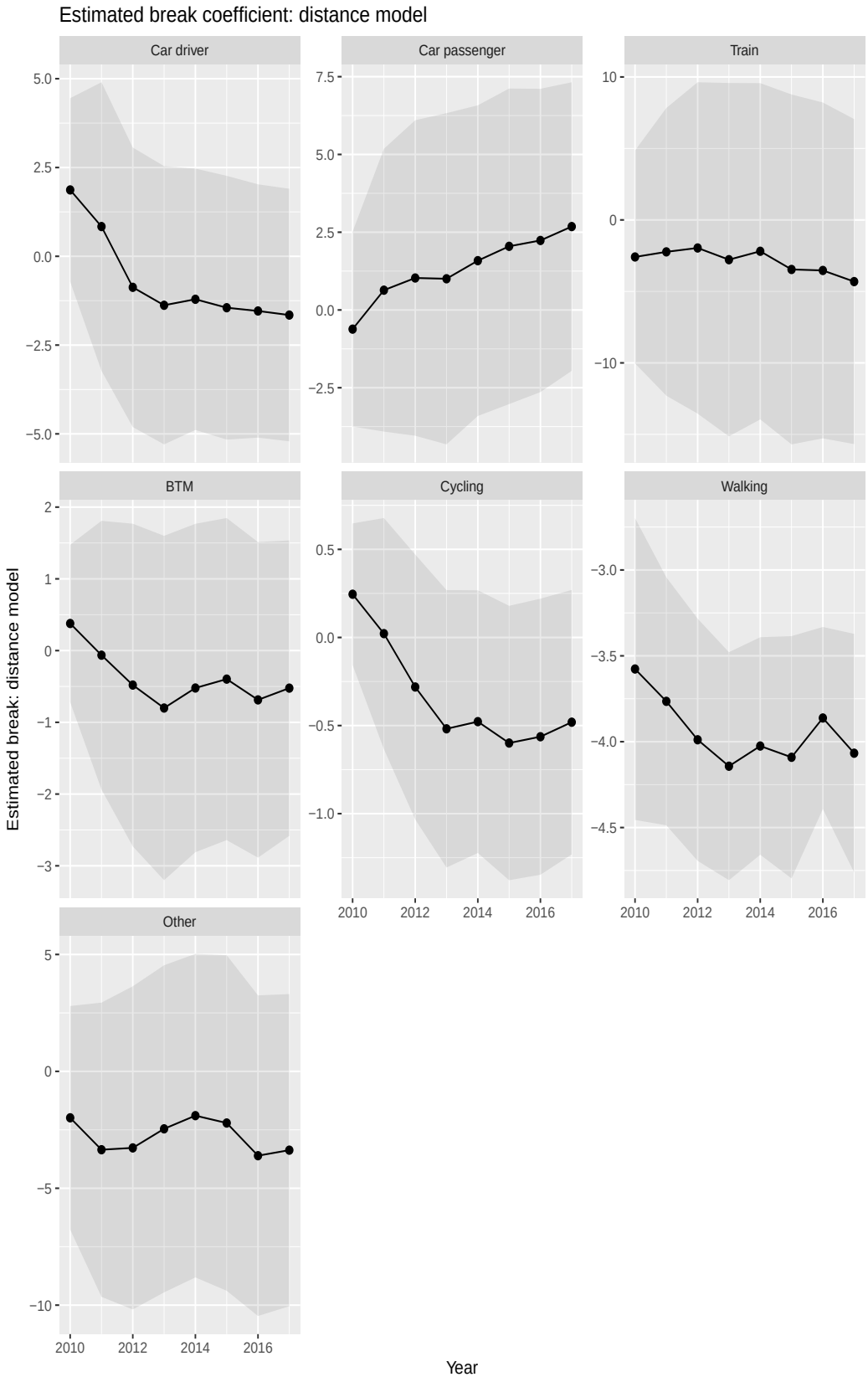

Figure A.217 In real time estimates from 2010 to 2017 for the discontinuity for number of trip legs at the mode level with approximate 95% confidence interval

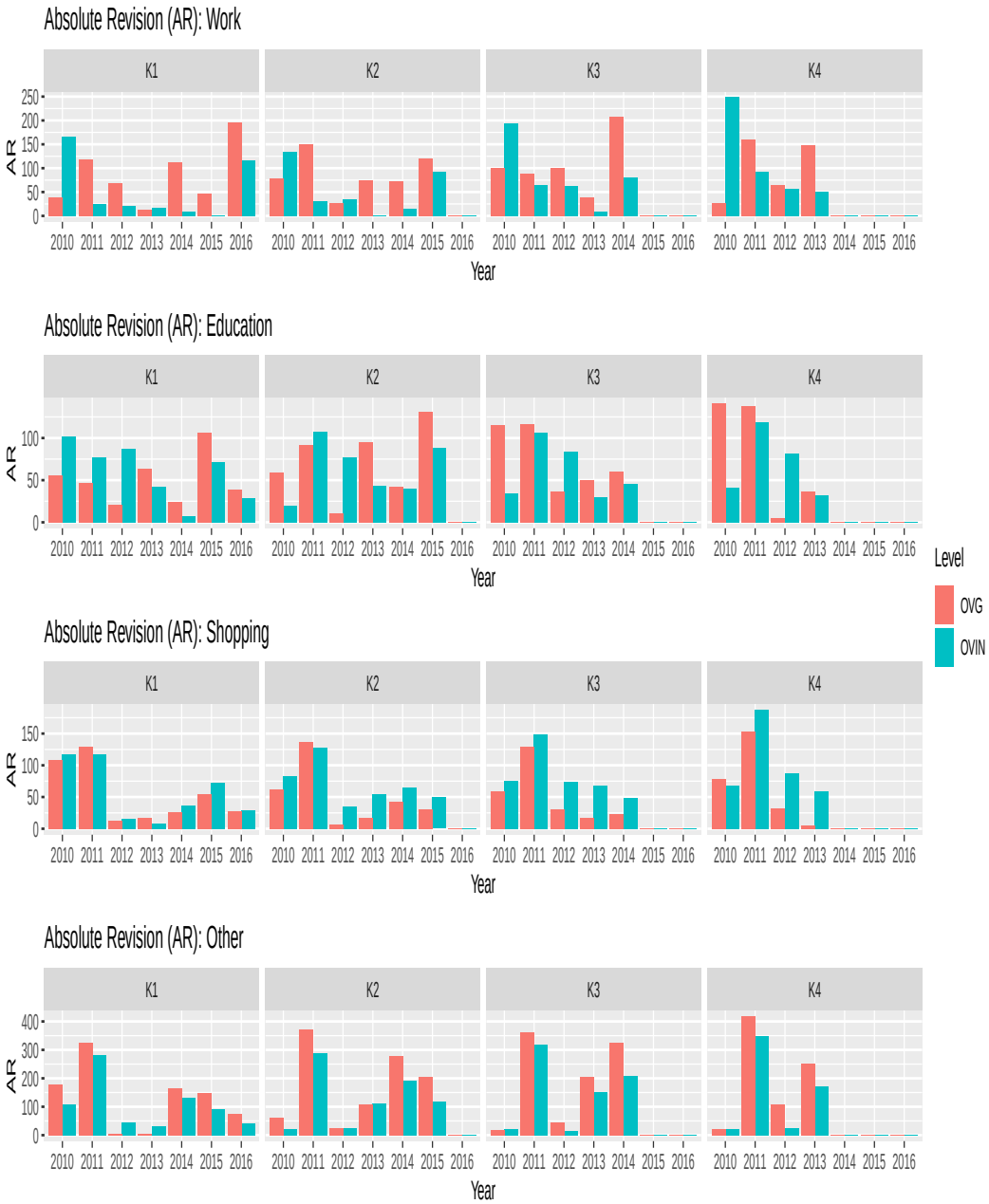

Figure A.218 Absolute Revision ($AR_k \times 100$) of the signals at overall level for mean distance per trip leg by purpose when the signals are published at OViN and OVG level for revision horizon $k = 1, 2, 3, 4$.

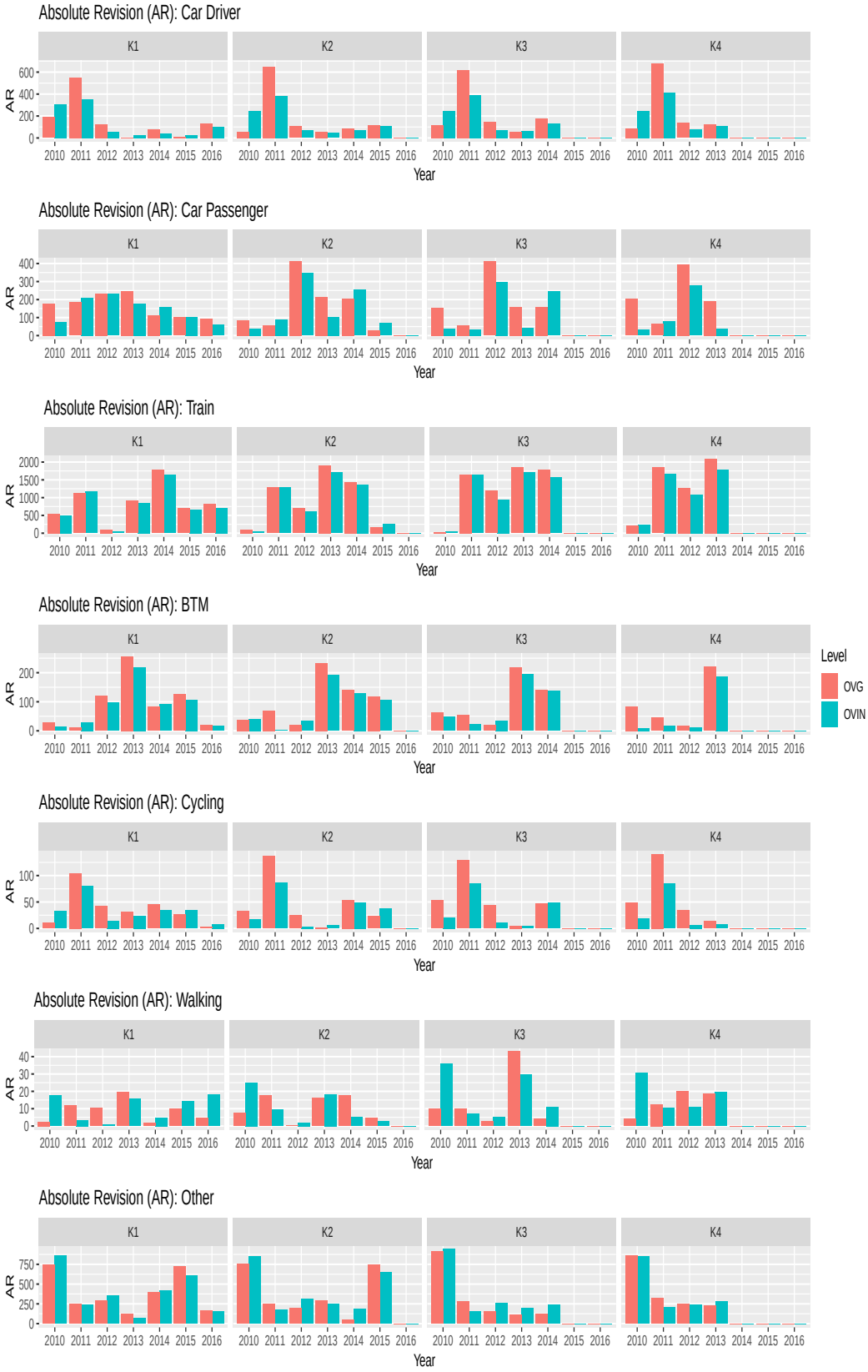

Figure A.219 Absolute Revision ($AR_k \times 100$) of the signals at overall level for mean distance per trip leg by mode when the signals are published at OViN and OVG level for revision horizon $k = 1, 2, 3, 4$.

	k=1		k=2		k=3		k=4	
	OVIN	OVG	OVIN	OVG	OVIN	OVG		
t	$AR_k \times 100$							
2010	142.2	129.1	75.0	32.7	85.8	12.4	76.9	31.5
2011	170.7	209.8	167.6	237.7	177.7	220.7	193.2	261.9
2012	14.0	18.9	1.4	6.0	19.4	54.3	26.2	75.2
2013	25.4	1.0	69.9	76.1	92.4	111.7	105.8	148.7
2014	81.9	111.7	116.9	159.0	130.7	198.0		
2015	55.5	66.4	76.0	113.6				
2016	23.2	52.9						
t	$RR_k\%$							
2010	1.41	1.27	0.74	0.32	0.85	0.12	0.76	0.31
2011	1.62	1.97	1.59	2.24	1.68	2.08	1.83	2.46
2012	0.13	0.18	0.01	0.06	0.19	0.52	0.25	0.72
2013	0.24	0.01	0.66	0.72	0.88	1.06	1.00	1.41
2014	0.78	1.06	1.11	1.50	1.24	1.87		
2015	0.52	0.63	0.72	1.07				
2016	0.22	0.50						

Table A.2 Absolute Revision ($AR_k \times 100$) and Relative Revision ($RR_k\%$) of the signals at overall level for mean distance per trip leg for revision horizon $k = 1, 2, 3, 4$ when the signals are published at OViN and OVG level.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information