This is an author produced version of Adjustment for time-invariant and time-varying confounders in 'unexplained residuals' (UR) models for longitudinal data within a causal framework, and associated challenges.

White Rose Research Online URL for this paper: http://eprints.whiterose.ac.uk/126041/

Article:
Arnold, KF orcid.org/0000-0002-0911-5029, Ellison, GTH, Gadd, SC et al. (4 more authors) (2018) Adjustment for time-invariant and time-varying confounders in 'unexplained residuals' (UR) models for longitudinal data within a causal framework, and associated challenges. Statistical Methods in Medical Research. ISSN 0962-2802 (In Press)

This is an author produced version of a paper accepted for publication in Statistical Methods in Medical Research. Uploaded in accordance with the publisher's self-archiving policy.
Adjustment for time-invariant and time-varying confounders in ‘unexplained residuals’ (UR) models for longitudinal data within a causal framework, and associated challenges

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Statistical Methods in Medical Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>SMM-17-0342.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>unexplained residuals model, conditional regression model, conditional analysis, conditional growth, conditional weight, conditional size, directed acyclic graph, DAG, causal inference, lifecourse epidemiology</td>
</tr>
<tr>
<td>Abstract:</td>
<td>‘Unexplained residuals’ (UR) models have been used within lifecourse epidemiology to model an exposure measured longitudinally at several time points in relation to a distal outcome. It has been claimed that these models have several advantages, including: the ability to estimate multiple total causal effects in a single model, and additional insight into the effect on the outcome of greater-than-expected increases in the exposure compared to traditional regression methods. We evaluate these properties, and prove mathematically how adjustment for confounding variables must be made within this modelling framework. Importantly, we explicitly place UR models in a causal framework using directed acyclic graphs (DAGs). This allows for theoretical justification of appropriate confounder adjustment, and provides a framework for extending our results to more complex scenarios than those examined in this paper. We also discuss several interpretational issues relating to UR models within a causal framework. We argue that UR models offer no additional insights compared to traditional regression methods, and, in fact, are more challenging to implement; moreover, they artificially reduce estimated standard errors. Consequently, we conclude that UR models, if used, must be implemented with great care.</td>
</tr>
</tbody>
</table>

https://mc.manuscriptcentral.com/smmr
Article type:

Original article

Corresponding author:

KF Arnold, Leeds Institute for Data Analytics, Worsley Building Level 11, University of Leeds, Leeds, LS2 9NL, United Kingdom; email: mt15kfa@leeds.ac.uk

Article title:

Adjustment for time-invariant and time-varying confounders in ‘unexplained residuals’ (UR) models for longitudinal data within a causal framework, and associated challenges

Authors:

KF Arnold\textsuperscript{a,b}, GTH Ellison\textsuperscript{a,b}, SC Gadd\textsuperscript{b}, J Textor\textsuperscript{c}, PWG Tennant\textsuperscript{a,d}, A Heppenstall\textsuperscript{a,e}, MS Gilthorpe\textsuperscript{a,b}

\textsuperscript{a}Leeds Institute for Data Analytics, University of Leeds, Leeds, United Kingdom
\textsuperscript{b}School of Medicine, University of Leeds, Leeds, United Kingdom
\textsuperscript{c}Tumor Immunology Lab, Radboud University Medical Centre, Nijmegen, The Netherlands
\textsuperscript{d}School of Healthcare, University of Leeds, Leeds, United Kingdom
\textsuperscript{e}School of Geography, University of Leeds, Leeds, United Kingdom

Abstract:

‘Unexplained residuals’ (UR) models have been used within lifecourse epidemiology to model an exposure measured longitudinally at several time points in relation to a distal outcome. It has been claimed that these models have several advantages, including: the ability to estimate multiple total causal effects in a single model, and additional insight into the effect on the outcome of greater-than-expected increases in the exposure compared to traditional regression methods. We evaluate these properties, and prove mathematically how adjustment for confounding variables must be made within this modelling framework. Importantly, we explicitly place UR models in a causal framework using directed acyclic graphs (DAGs). This
allows for theoretical justification of appropriate confounder adjustment, and provides a framework for extending our results to more complex scenarios than those examined in this paper. We also discuss several interpretational issues relating to UR models within a causal framework. We argue that UR models offer no additional insights compared to traditional regression methods, and, in fact, are more challenging to implement; moreover, they artificially reduce estimated standard errors. Consequently, we conclude that UR models, if used, must be implemented with great care.

**Key words:**

unexplained residuals model, conditional regression model, conditional analysis, conditional growth, conditional weight, conditional size, directed acyclic graph, DAG, causal inference, lifecourse epidemiology
1. Background

Within the field of lifecourse epidemiology, there is substantial interest in modelling the relationship between an exposure \( x \) measured longitudinally at several time points (i.e. \( x_1, x_2, \ldots, x_k \)) and a subsequent outcome \( y \) measured once later in life (hereafter referred to as a distal outcome); such a relationship can be helpfully summarised in Figure 1a in the form of a directed acyclic graph (DAG)\(^1\).

DAGs are pictorial representations of hypothesised causal relationships between variables in which: variables (nodes) are connected via unidirectional arrows (directed edges), which represent direct causal relationships; and no directed loops (i.e. circular paths) between variables are permitted. Nodes may be either: endogenous, having at least one causally preceding variable represented in the graph; or exogenous, having none\(^2\). All unexplained causes of the endogenous nodes \( x_2, \ldots, x_k, y \) in Figure 1a are represented by the variables \( e_{x1}, \ldots, e_{xk}, e_y \), respectively. Whilst there are many useful applications for DAGs in epidemiologic research, perhaps the most beneficial is their ability to identify suitable sets of covariates for removing bias due to confounding between an exposure and outcome\(^3,4\), which occurs whenever both variables share one or more common causes. For this reason, DAGs are increasingly being used in epidemiology, as they provide a framework for estimating the total causal effect of an exposure on an outcome\(^4\).

[Insert Figure 1a-c]

Using a causal framework to (correctly) model the scenario in Figure 1a may also have additional utility in identifying and quantifying important periods of change in the exposure that are causally related to the outcome. However, one challenge to such applications is that successive measurements of an exposure over time may be highly correlated with one another and therefore likely to suffer collinearity.
when analysed in relation to a distal outcome. Consequently, there has been extensive debate regarding the best way to model these types of longitudinally measured variables; a recent review\(^5\) of analytical and modelling techniques has identified a range of different approaches, including z-score plots, regression with change scores, multilevel and latent growth curve models, and growth mixture models. Nonetheless, one of the most straightforward methods in use is a series of standard multivariable regression models.

1.1. Standard regression method

When using this approach, each longitudinal measurement of the exposure variable is treated as a separate entity that is subject to confounding by all previous measurements of that variable – the total number of models needed therefore being equal to the total number of time points at which the exposure has been measured.

As an example, the simplest scenario would involve just two measurements of the exposure \(x\) (i.e. \(x_1\) and \(x_2\), measured at time points 1 and 2, respectively), and a distal outcome, \(y\), where all variables are continuous in nature. Here, two standard regression models (denoted \(\hat{y}_S^{(i)}\), for \(i = 1,2\)) would need to be constructed to estimate the total causal effect of each distinct measurement of \(x\) on \(y\), i.e.

\[
\hat{y}_S^{(1)} = \hat{a}_0^{(1)} + \hat{a}_{x_1}^{(1)} x_1 \\
\hat{y}_S^{(2)} = \hat{a}_0^{(2)} + \hat{a}_{x_1}^{(2)} x_1 + \hat{a}_{x_2}^{(2)} x_2.
\]

(Eq.1)

(Eq.2)

Importantly, to estimate the total causal effect of \(x_1\) on \(y\) in Eq.1, adjustment for \(x_2\) is inappropriate, as it lies on the causal path between \(x_1\) and \(y\) (i.e. \(x_2\) is a mediator); in fact, adjustment for \(x_2\) might invoke bias in the causal interpretation due to a phenomenon known as the ‘reversal paradox’\(^5-7\). In contrast, to
estimate the total causal effect of \( x_2 \) on \( y \) in Eq.2, adjustment for \( x_1 \) is appropriate, since it confounds the desired relationship (i.e. \( x_1 \) causally precedes both \( x_2 \) and \( y \), potentially creating a spurious relationship between them). For this reason, in either model, it is only possible to interpret the coefficient of the last/most recent measurement of \( x \) (the exposure) as a total causal effect\(^1\), which encompasses all direct and indirect causal pathways between an exposure and outcome. No such interpretation is possible (nor should it be attempted) for the coefficient of the earlier measurement of \( x \) in Eq.2, as it operates purely as a confounder.

### 1.2. Unexplained residuals (UR) method

To circumvent the need for multiple models, Keijzer-Veen\(^8\) has suggested an alternative approach that would combine the information contained within each of the two separate models (Eq.1 and Eq.2) into a single composite regression model using ‘unexplained residuals’. As originally proposed\(^9\), such a model allows the researcher to quantify the total effects of both the initial measurement of \( x \) (i.e. \( x_1 \)) and subsequent change in \( x \) on the outcome \( y \). The proposed approach contains two steps but is straightforward in principle.

First, the most recent measurement of \( x \) (i.e. \( x_2 \)) is regressed on the earlier measurement of \( x \) (i.e. \( x_1 \)):

\[
    x_2 = \hat{\varphi}_0^{(2)} + \hat{\varphi}_{x1}^{(2)} x_1 + e_{x2}.
\]

(Eq.3)

This produces a measure of each observation’s ‘expected’ value of \( x_2 \) as predicted by its value of \( x_1 \). The difference between the expected value of \( x_2 \) (i.e. \( \hat{\varphi}_0^{(2)} + \hat{\varphi}_{x1}^{(2)} x_1 \)) and the observed value of \( x_2 \) amounts to the residual term \( e_{x2} \). Put another way, \( e_{x2} \) represents the part of \( x_2 \) ‘unexplained’ by \( x_1 \).

Second, \( y \) is regressed on both the initial exposure \( x_1 \) and subsequent residual term \( e_{x2} \):
\[ \hat{y}^{(2)}_{UR} = \hat{y}^{(2)}_{S} + \hat{y}^{(2)}_{x1} + \hat{y}^{(2)}_{ex2}. \]  
(Eq.4)

According to Keijzer-Veen et al.\(^9\), the key advantages of conducting regression using the composite ‘unexplained residuals’ (UR) model (Eq.4) are that:

1. The UR model produces the same estimated outcome values as the standard regression model in Eq.2 (i.e. \(\hat{y}^{(2)}_{S} = \hat{y}^{(2)}_{UR}\));

2. The estimated total effect sizes (i.e. the coefficient values) produced by individual standard regression models (Eq.1 and Eq.2) are equal to those estimated within the UR model (i.e. \(\hat{\alpha}^{(1)}_{x1} = \hat{\alpha}^{(2)}_{x1}\) and \(\hat{\alpha}^{(2)}_{x2} = \hat{\alpha}^{(2)}_{ex2}\)); thus, multiple coefficients in a single model may be interpreted;

3. The UR model provides additional insight (via the coefficient \(\hat{\alpha}^{(2)}_{ex2}\) in Eq.4) into the effect of \(x\) increasing more than expected upon \(y\); and

4. The initial exposure \(x_1\) and subsequent residual term \(e_{x2}\) are mathematically independent (i.e. orthogonal).

Succinctly, the two models \(\hat{y}^{(2)}_{S}\) and \(\hat{y}^{(2)}_{UR}\) are algebraically equivalent, but \(\hat{y}^{(2)}_{UR}\) makes interpretation of the separate influence of the initial measurement of the exposure \(x\) (i.e. \(x_1\)) and subsequent changes in \(x\) more straightforward than do (multiple) standard regression models \(\hat{y}^{(1)}_{S}\) and \(\hat{y}^{(2)}_{S}\).

Within the epidemiological literature, UR models have been used under a number of different names. In addition to ‘regression with unexplained residuals’ (as first proposed by Keijzer-Veen et al.\(^9\)-\(^11\)), other studies have referred to: ‘unexplained residual regression’\(^12\); ‘method of unexplained residuals’\(^13\); ‘conditional linear regression’\(^12\); ‘conditional (regression) models’\(^5\), \(^14\); ‘regression with conditional growth measures’\(^14\); ‘conditional growth models’\(^15\)-\(^18\); ‘conditional weight models’\(^19\); and ‘conditional...
The terms ‘conditional growth’ and ‘conditional size’ – and additional variations thereof – are also commonly used to refer to the difference between observed and expected size measurements\(^{5, 15, 18, 25-39}\). To avoid further confusion, the residual term representing the difference between the observed and expected values of an exposure produced in the manner proposed by Keijzer-Veen et al. (as in Eq.3) will be henceforth referred to as the ‘unexplained residuals (UR) term’, and the models themselves (as in Eq.4) will be referred to as ‘unexplained residuals (UR) models’.

Despite the numerous names given to these models, the process remains essentially the same as that first proposed. Indeed, several authors have extended the original model to examine scenarios involving several measurements of an exposure \(x\) (i.e. \(x_1, x_2, \ldots, x_k\)); UR models in these extended applications thus include several UR terms\(^{5, 12, 13, 16-41}\). In general, each UR term \(e_{xi}\) is derived from the regression of each measured value \(x_i\) on all previous measurements \(x_1, x_2, \ldots, x_{i-1}\), for \(2 \leq i \leq k\)\(^{12, 16, 18, 22, 24, 25, 27, 29, 31-34, 36, 39, 40}\), though some researchers have deviated from this procedure\(^{13, 26, 35, 37, 41}\); the outcome \(y\) is then regressed on \(x_1\) and all subsequent UR terms \(e_{x2}, e_{x3}, \ldots, e_{xk}\).

Many researchers have further extended the original UR models by adjusting for additional confounding variables (i.e. over and above the confounding of prior measurements of the exposure), though there is, as yet, little consensus as to whether or how such adjustments should be performed. For example, Horta et al.\(^{16}\) made no adjustments for potential confounders when deriving their UR terms, but did make adjustments within their composite UR model. In contrast, Gandhi et al.\(^{18}\) adjusted for just one potential confounder (gender) when creating their UR terms, but also made further adjustments to the composite UR model (for gender and other variables). Adair et al.\(^{25}\) created their UR terms using site- and sex-stratified linear regressions that were also adjusted for age, and made further adjustments for age, sex, and study site in their subsequent composite UR models. Indeed, there are many other
examples of different approaches to confounder adjustment, but none of these have been adequately
and explicitly justified by the researchers concerned, even though it appears that they did so in order to
make causal inferences.

2. Research aims

The potential impact of using alternative approaches to adjust for confounding when constructing and
using UR terms has yet to be fully evaluated. Indeed, Keijzer-Veen et al. did not address confounding
variables in their original paper, and there has been little to no discussion (or analysis) of this issue by
subsequent authors using this approach. It therefore remains unclear whether UR models that include
confounders offer the same purported benefits as those lacking (or ignoring) confounders, and there is
no clear indication of how potential confounders should be treated by analyses using these models. This
is an issue of particular relevance to researchers seeking to infer causality from individual coefficient
estimates, since inappropriate adjustment for covariates (which includes both the failure to adjust for
genuine confounders and the adjustment for mediators mistaken for confounders) can lead to biased
causal inferences. For this reason, UR models are likely to have limited practical utility unless they are
able to accommodate confounding variables appropriately. The fact that UR models have not been
developed or analysed within a causal framework also creates uncertainty about their utility for making
causal inferences.

Therefore, the aims of the present study were to: (1) confirm that the approach proposed by Keijzer-
Veen et al. may be extended to a scenario involving $k$ longitudinal measurements of an exposure $x$ in
the absence of any additional confounding; (2) determine whether it is possible (and if so, how might it
be possible) to adjust for additional confounders within the UR modelling framework; (3) evaluate the
benefits of UR models claimed by Keijzer-Veen et al.; and (4) offer recommendations for future use of UR models. The present study examines two very different types of potential confounders: time-invariant (which require/provide measurements taken at a single time point and remain constant across the life course, e.g. sex); and time-varying (for which measurements are collected at multiple time points across the life course – usually concurrent to measurements of the exposure – because the value of the variable may change, e.g. socioeconomic position).

These aims are summarised in the DAGs presented in Figures 1a, 2a, and 3a, which depict three general scenarios drawn from lifecourse epidemiology, each of which will be examined in the analyses that follow. Each DAG relates \( k \) longitudinally measured exposure variables \( x_1, x_2, \ldots, x_k \) (i.e. \( x \) measured at time points \( 1, 2, \ldots, k \)) to a distal outcome \( y \) (measured at some point either concurrent to or following \( k \)) under three very different circumstances: (1a) in the absence of any additional confounders; (2a) in the presence of an additional time-invariant confounder \( m \); and (3a) in the presence of an additional time-varying confounder \( m_1, m_2, \ldots, m_k \). All DAGs are drawn forwardly saturated (i.e. where each node may causally affect all future nodes), and all unexplained causes of endogenous nodes are represented by the variable \( e \) and depicted as independent (i.e. we assume no unobserved confounding). The explicit inclusion of these three DAGs in Figures 1a, 2a, and 3a is intended not only to visually illustrate each of the scenarios that will be examined, but also, importantly, to situate the analyses that follow within a causal framework.

[Insert Figure 2a-c]

[Insert Figure 3a-c]

Sections 3 through 9, which follow, provide: the three key properties of UR models that will be evaluated for the scenarios in Figures 1a, 2a, and 3a (§3); DAG-based and mathematical examinations of
the UR models for the scenarios given in Figure 1a (§4), 2a (§5), and 3a (§6); a discussion of several interpretational issues that arise for UR models when placed within a causal framework, including an evaluation of the claim that UR models provide greater insight than standard regression methods (§7); an argument outlining how UR models produce artificially reduced standard errors and how this might be corrected (§8); and recommendations for future use and interpretation of UR models, particularly as these relate to the inclusion of confounders (§9).

3. Key properties of UR models

In the following sections, we evaluate the mathematical properties of the original UR models after extending them to include \( k \) measurements of a continuous exposure \( x \): in the absence of any additional confounding (§4); in the presence of a single additional time-invariant confounder \( m \) (§5); and in the presence of a single additional time-varying confounder with sequential values \( m_1, m_2, \ldots, m_k \) (§6). These properties are:

- **Property (i):** The outcome values predicted by the final standard regression model (for the final measurement of the exposure variable, \( x_k \)) are equal to those predicted by the composite UR model.

- **Property (ii):** The estimated coefficient for \( x_1 \) in the initial standard regression model (for the first measurement of the exposure variable, \( x_1 \)) is equal to the estimated coefficient for \( x_1 \) in the composite UR model.

- **Property (iii):** The estimated coefficient for each \( x_i \) in its individual standard regression model (i.e. for designated exposure \( x_i \)) is equal to the estimated coefficient for the corresponding UR term \( e_{xi} \) in the composite UR model.
From a causal inference perspective, only Properties (ii) and (iii) are meaningful, since the focus is on individual coefficient estimates as opposed to predicted outcomes. Nevertheless, we evaluate all three properties in Sections 4 through 6, and leave discussion of interpretational issues until later in the paper (§8).

4. UR models: No confounders (Figure 1a)

Before considering any additional confounding variables, we first consider the straightforward scenario depicted in Figure 1a. We provide: definitions of the standard regression models, UR terms, and UR models (§4.1); an analysis of UR models within a causal framework (§4.2); and arguments for why Properties (i) – (iii) are upheld (§4.3).

4.1. Definitions

We define the ordinary least-squares (OLS) regression model $\hat{y}_S^{(i)}$ for estimating the total causal effect of each measurement of the exposure variable $x_i$ (for $1 \leq i \leq k$) on $y$ as:

$$\hat{y}_S^{(i)} = \hat{a}_0^{(i)} + \hat{a}_{x1}^{(i)}x_1 + \hat{a}_{x2}^{(i)}x_2 + \cdots + \hat{a}_{xi}^{(i)}x_i.$$  (Eq.5)

A visual depiction of Eq.5 is given in Figure 1b. Because the relationship between each $x_i$ and $y$ is confounded by all previous measurements of $x$ (i.e. $x_1, ..., x_{i-1}$), these covariates must be adjusted for. However, as discussed in Section 1, only the coefficient of the last/most recent measurement of $x$ (i.e. $\hat{a}_{xi}^{(i)}$) may be interpreted as a total causal effect.

To create UR terms according to the process established by Keijzer-Veen et al.⁹, each measurement of the exposure $x_i$ is regressed on all previous measurements of $x$ (for $2 \leq i \leq k$):
\[ x_i = y_0^{(i)} + \beta_1^{(i)} x_1 + \beta_2^{(i)} x_2 + \cdots + \beta_{i-1}^{(i)} x_{i-1} + e_{xi}. \]  

(Eq.6)

The UR term \( e_{xi} \) thus represents the difference between the actual value of \( x_i \) and the value of \( x_i \) as predicted by all previous measurements of \( x \).

Lastly, we define the UR model \( \hat{y}^{(i)}_{UR} \) (for \( 1 \leq i \leq k \)), which represents the outcome \( y \) as function of the initial value of the exposure \( x_1 \) and subsequent ‘unexplained’ increases \( e_{x2}, \ldots, e_{xi} \):

\[ \hat{y}^{(i)}_{UR} = \lambda_0^{(i)} + \lambda_1^{(i)} x_1 + \lambda_2^{(i)} e_{x2} + \cdots + \lambda_{i-1}^{(i)} e_{xi}. \]  

(Eq.7)

The composite UR model \( \hat{y}^{(k)}_{UR} \) thus represents the outcome \( y \) as function of the initial value of the exposure \( x_1 \) and all subsequent ‘unexplained’ increases \( e_{x2}, \ldots, e_{xk} \). The UR modelling process is summarised in Figure 1c, depicting \( k - 1 \) regressions of \( x_i \) on \( x_1, \ldots, x_{i-1} \) (Eq.6) and one composite UR regression model (Eq.7, with \( i = k \)).

### 4.2. A causal framework

Within the causal framework provided by Figure 1a, the unique properties of UR models can be visualised. If we were naively to model \( x_1, x_2, \ldots, x_k \) simultaneously, only the coefficient of the final measurement \( x_k \) could be interpreted as a total causal effect on \( y \); the coefficients of \( x_1, \ldots, x_{k-1} \) would represent only the direct effects of each measurement on \( y \), because all future measurements would fully mediate the respective relationship and all backdoor paths would be blocked by preceding measurements. However, by modelling \( x_1, e_{x2}, \ldots, e_{xk} \) (as in a UR model), we encounter no mediation problems due to the fact that, by construction, the UR terms remain wholly independent of the other terms in the model. In fact, by placing the UR model in a causal framework, we are able to see that the
UR terms $e_{x_2}, \ldots, e_{x_k}$ are essentially instrumental variables (IVs) for $x_2, \ldots, x_k$, respectively, which have been produced by the modelling process.\footnote{The process has similarities with the two-stage least squares regression method, a form of instrumental variable analysis commonly encountered in economics research.}

All techniques based on linear regression, including UR models, assume that the causal relationships between variables are linear functions. If that is the case, we may parameterise a DAG (as in Figure 1a) by assigning a single coefficient to every arrow and assuming all variables to have a variance of one. The method of path coefficients then allows us to determine the 'true' total causal effects in the data generating process. Take $x_2$ as an example, where $k = 3$. The total effect of $x_2$ on $y$ encompasses the direct effect from $x_2 \rightarrow y$ and all indirect effects (of which there is only one in this scenario): $x_2 \rightarrow x_3 \rightarrow y$. We introduce the notation $p_{a \rightarrow b}$ to represent the coefficient of the arrow $a \rightarrow b$. Table 1 gives the total effects of $x_2$ on $y$ and of $e_{x_2}$ on $y$, with both total effects decomposed into their respective direct and indirect effects. From Table 1, we see that the total effect of $x_2$ on $y$ is equal to the total effect of $e_{x_2}$ on $y$; this is because there are no direct paths between $e_{x_2}$ and $y$, and all indirect paths pass through $x_2$ (with $p_{x_2 e_{x_2}}$ being equal to one, as in Figure 1c).
For Peer Review

### 4.3. Covariate orthogonality and Properties (i) – (iii)

In addition to the graph-based approach in the preceding section, we are able to prove mathematically that Properties (i) – (iii) are upheld for the scenario given in Figures 1a-b. In summary, these properties are:

- Property (i): \( \tilde{y}_s^{(k)} = \tilde{y}_{UR}^{(k)} \)
- Property (ii): \( \tilde{\alpha}^{(i)}_{x_1} = \tilde{\lambda}_{x_1}^{(k)} \)
- Property (iii): \( \tilde{\alpha}^{(i)}_{x_1} = \tilde{\lambda}_{x_1}^{(k)} \)

Eq.5 – Eq.7 are summarised in Table 2; the standard regression models \( \tilde{y}_s^{(i)} \) (for \( 1 \leq i \leq k \)) and composite UR model \( \tilde{y}_{UR}^{(k)} \) (in which the UR terms have been produced via the regression of each measurement of \( x \) on all previous measurements, as in Eq.5) contained therein are guaranteed to satisfy Properties (i) – (iii). These properties of UR models rely crucially on all UR terms \( e_{x_2}, ..., e_{x_k} \) being orthogonal to all other covariates in the composite UR model \( \tilde{y}_{UR}^{(k)} \).

### Table 1: Total effect of \( x_2 \) on \( y \) estimated by a standard regression model compared to total effect of \( e_{x_2} \) on \( y \) estimated by an equivalent UR model (Figure 1a, with \( k = 3 \)).

<table>
<thead>
<tr>
<th>Exposure:</th>
<th>Path:</th>
<th>Effect size:</th>
<th>Total effect:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_2 )</td>
<td>Direct: ( x_2 \rightarrow y )</td>
<td>( p_{yx_2} )</td>
<td>( p_{yx_2} + p_{yx_2} \cdot p_{yx_2} )</td>
</tr>
<tr>
<td></td>
<td>Indirect: ( x_2 \rightarrow x_3 \rightarrow y )</td>
<td>( p_{x_2x_2} \cdot p_{yx_2} )</td>
<td>( p_{yx_2} + p_{x_2x_2} \cdot p_{yx_2} )</td>
</tr>
<tr>
<td>( e_{x_2} )</td>
<td>Direct: n/a</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Indirect: ( e_{x_2} \rightarrow x_2 \rightarrow y )</td>
<td>( p_{x_2e_{x_2}} \cdot p_{yx_2} )</td>
<td>( p_{yx_2} + p_{x_2e_{x_2}} \cdot p_{yx_2} )</td>
</tr>
<tr>
<td></td>
<td>( e_{x_2} \rightarrow x_2 \rightarrow x_3 \rightarrow y )</td>
<td>( p_{x_2e_{x_2}} \cdot p_{x_2x_2} \cdot p_{yx_2} )</td>
<td>( p_{yx_2} + p_{x_2e_{x_2}} \cdot p_{x_2x_2} \cdot p_{yx_2} )</td>
</tr>
</tbody>
</table>
We illustrate this property, and explain how it is exploited to ensure Properties (i) – (iii) are upheld.

Formal proofs are provided in the attached Appendix 1.

<table>
<thead>
<tr>
<th>$i$ = 1:</th>
<th>Standard regression model $y^{(1)}_S$:</th>
<th>UR model $y^{(1)}_{UR}$:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\alpha}<em>0^{(1)} + \alpha</em>{x1}^{(1)} x_1$</td>
<td>$\hat{\lambda}^{(1)}<em>0 + \lambda</em>{x1}^{(1)} x_1$</td>
</tr>
<tr>
<td>$i$ = 2:</td>
<td>$\hat{\alpha}<em>0^{(2)} + \alpha</em>{x1}^{(2)} x_1 + \alpha_{x2}^{(2)} x_2$</td>
<td>$\hat{\lambda}<em>0^{(2)} + \lambda</em>{x1}^{(2)} x_1 + \lambda_{x2}^{(2)} e_{x2}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$i$ = $k$:</td>
<td>$\hat{\alpha}<em>0^{(k)} + \alpha</em>{x1}^{(k)} x_1 + \alpha_{x2}^{(k)} x_2 + \ldots + \alpha_{xk}^{(k)} x_k$</td>
<td>$\hat{\lambda}<em>0^{(k)} + \lambda</em>{x1}^{(k)} x_1 + \lambda_{x2}^{(k)} e_{x2} + \ldots + \lambda_{xk}^{(k)} e_{xk}$</td>
</tr>
</tbody>
</table>

Table 2: For the scenario depicted in Figures 1a-b, the standard regression model $y^{(i)}_S$ necessary for estimating the total causal effect of each exposure $x_i$ on $y$, and the corresponding UR model $y^{(i)}_{UR}$ for $1 \leq i \leq k$.

In Table 2, note that each regression model (for both the standard and UR methods) contains one more covariate than the model preceding it. In the column of standard regression models, each row contains an additional $x_i$ term; in the column of UR models, each row contains an additional $e_{xi}$ term.

Typically, the inclusion of an additional covariate in a regression model changes the coefficient(s) estimated for other covariates because their covariance would be nonzero. For example, the addition of $x_2$ in $y^{(2)}_S$ will undoubtedly change the estimated coefficient for $x_1$ in $y^{(2)}_S$ compared to $y^{(1)}_S$, because $x_1$ and $x_2$ are two measurements of the same variable and thus will have a nonzero covariance (i.e. correlation ≠ 0). This nonzero covariance is what is exploited by adjustment for confounders – if two covariates did not covary, then adjustment would not be necessary in the first place.

However, a UR model upholds Properties (ii) and (iii) specifically because its covariates do not covary.

The addition of $e_{x2}$ in $y^{(2)}_{UR}$ does not change the estimated coefficient for $x_1$ in $y^{(2)}_{UR}$ compared to $y^{(1)}_r$ because $x_1$ and $e_{x2}$ are orthogonal (i.e. correlation = 0). This orthogonality is ensured as an artefact of OLS regression; because $e_{x2}$ represents the residual term from the regression of $x_2$ on $x_1$ by definition (Eq.6), it is guaranteed to be orthogonal to $x_1$. 
In fact, it can easily be shown that all UR terms $e_{x_2}, \ldots, e_{x_k}$ are orthogonal to one another by construction. For any UR term $e_{x_i}$, it holds that $e_{x_i}$ is orthogonal to $x_1, \ldots, x_{i-1}$. Because preceding UR terms $e_{x_2}, \ldots, e_{x(i-1)}$ are themselves linear combinations of $x_1, \ldots, x_{i-1}$ (Eq.6), it follows that $e_{x_i}$ is orthogonal to $e_{x_2}, \ldots, e_{x(i-1)}$, for $2 \leq i \leq k$. Using this information, we can easily conclude that the addition of subsequent UR terms in the set of UR models in Table 2 will leave the coefficients of all other covariates unchanged. Thus, it only remains to be shown that the estimated coefficients for $x_1$ and the UR terms $e_{x_2}, \ldots, e_{x_k}$ are themselves equivalent to the coefficients for $x_1, x_2, \ldots, x_k$ as estimated in their individual standard regression models, respectively.

Property (i):

First, it must be noted that each UR model is nothing more than a reparameterisation of the corresponding standard regression model (i.e. $\hat{\beta}^{(i)}_S = \hat{\beta}^{(i)}_{UR}$ for each row in Table 2). Each standard regression model $\hat{\beta}^{(i)}_S$ represents $y$ as a function of $x_1, \ldots, x_i$. In contrast, each UR model $\hat{\beta}^{(i)}_{UR}$ represents $y$ as a function of $x_1, e_{x_2}, \ldots, e_{x_i}$. However, $e_{x_i}$ is itself a function of $x_1, \ldots, x_i$ (Eq.5), and thus it follows that the UR model $\hat{\beta}^{(i)}_{UR}$ itself is also a function of $x_1, \ldots, x_i$. Because $\hat{\beta}^{(i)}_S$ and $\hat{\beta}^{(i)}_{UR}$ are both functions of the same covariates, it follows that $\hat{\beta}^{(k)}_S = \hat{\beta}^{(k)}_{UR}$, thereby satisfying Property (i).

Property (ii):

It is trivially true that the coefficients estimated for $x_1$ in the first standard regression model $\hat{\beta}^{(1)}_S$ and corresponding UR model $\hat{\beta}^{(1)}_{UR}$ will be equal (i.e. $\hat{\beta}^{(1)}_S = \hat{\beta}^{(1)}_{UR}$) because the models are themselves equivalent. All subsequent UR terms $e_{x_2}, \ldots, e_{x_k}$ are orthogonal to $x_1$ and to one another; therefore, it follows that the estimated coefficient of $x_1$ will be equivalent for all UR models in Table 1 (i.e.
\[ \hat{\lambda}_{x_1}^{(1)} = \hat{\lambda}_{x_1}^{(2)} = \cdots = \hat{\lambda}_{x_1}^{(k)} \]. This ensures that the coefficient of \( x_1 \) in \( \hat{y}_{s}^{(1)} \) (which represents the total effect of \( x_1 \) on \( y \)) will be unchanged in the composite UR model \( \hat{y}_{UR}^{(k)} \) (i.e. \( \hat{\lambda}_{x_1}^{(1)} = \hat{\lambda}_{x_1}^{(k)} \)).

**Property (iii):**

Lastly, we can show that the coefficient for \( e_{x_1} \) (i.e. \( \hat{\lambda}_{e_{x_1}}^{(i)} \)) in a UR model is equal to the estimated total effect of \( x_i \) (i.e. \( \hat{\lambda}_{x_i}^{(i)} \)) in the corresponding standard regression model. To this end, we consider the following standard regression and corresponding UR models, respectively:

\[
\hat{y}_{s}^{(i)} = \hat{\lambda}_{0}^{(i)} + \hat{\lambda}_{x_1}^{(i)} x_1 + \hat{\lambda}_{x_2}^{(i)} x_2 + \cdots + \hat{\lambda}_{x_i}^{(i)} x_i
\]

\[
\hat{y}_{UR}^{(i)} = \hat{\lambda}_{0}^{(i)} + \hat{\lambda}_{x_1}^{(i)} x_1 + \hat{\lambda}_{x_2}^{(i)} e_{x_2} + \cdots + \hat{\lambda}_{e_{x_1}}^{(i)} e_{x_1} .
\]

We may set these two equations equal to one another (due to Property (i)), substitute the expansions for \( e_{x_2}, \ldots, e_{x_i} \) (Eq.5) into the UR model and rearrange, thereby producing:

\[
\hat{\lambda}_{0}^{(i)} + \hat{\lambda}_{x_1}^{(i)} x_1 + \hat{\lambda}_{x_2}^{(i)} x_2 + \cdots + \hat{\lambda}_{x_i}^{(i)} x_i = \hat{\lambda}_{0}^{(i)} + \hat{\lambda}_{x_1}^{(i)} x_1 + \hat{\lambda}_{x_2}^{(i)} e_{x_2} + \cdots + \hat{\lambda}_{e_{x_1}}^{(i)} e_{x_1}
\]

\[
= \hat{\lambda}_{0}^{(i)} + \hat{\lambda}_{x_1}^{(i)} x_1 + \hat{\lambda}_{x_2}^{(i)} [-\hat{y}_{0}^{(i)} - \hat{\lambda}_{x_1}^{(i)} x_1 + \hat{y}_{x_1}^{(i)} x_1 + \cdots]
\]

\[
= \hat{\lambda}_{e_{x_1}}^{(i)} [-\hat{y}_{0}^{(i)} - \hat{y}_{x_1}^{(i)} x_1 - \hat{y}_{x_2}^{(i)} x_2 - \cdots - \hat{y}_{x_{i-1}}^{(i)} x_{i-1} + \hat{y}_{x_i}^{(i)} x_i]
\]

\[
= [\hat{\lambda}_{0}^{(i)} - \hat{\lambda}_{e_{x_1}}^{(i)} \hat{y}_{0}^{(i)} - \cdots - \hat{\lambda}_{e_{x_1}}^{(i)} \hat{y}_{x_{i-1}}^{(i)}] +
\]

\[
[\hat{\lambda}_{x_1}^{(i)} - \hat{\lambda}_{e_{x_2}}^{(i)} \hat{y}_{x_1}^{(i)} - \cdots - \hat{\lambda}_{e_{x_2}}^{(i)} \hat{y}_{x_{i-1}}^{(i)}] x_1 +
\]

\[
[\hat{\lambda}_{x_2}^{(i)} - \hat{\lambda}_{e_{x_3}}^{(i)} \hat{y}_{x_2}^{(i)} - \cdots - \hat{\lambda}_{e_{x_3}}^{(i)} \hat{y}_{x_{i-1}}^{(i)}] x_2 + \cdots +
\]

\[
[\hat{\lambda}_{e_{x_i}}^{(i)}] x_i .
\]  

(Eq.8)
From Eq.8 above, it becomes clear that the coefficients for \( x_i \) in \( y^{(i)}_S \) and \( e_{xi} \) in \( y^{(i)}_{UR} \) are equal (i.e. \( \hat{\alpha}_{xi}^{(i)} = \lambda_{e_{xi}}^{(i)} \)). Again, we invoke the property of orthogonality to conclude that the estimated coefficient for \( e_{xi} \) will be equivalent for all UR models in Table 2 (i.e. \( \hat{\lambda}_{ext}^{(1)} = \hat{\lambda}_{ext}^{(2)} = \cdots = \hat{\lambda}_{ext}^{(k)} \)). This ensures that the coefficient of \( e_{xi} \) in \( y^{(k)}_S \) (which represents the total effect of \( x_i \) on \( y \)) will be unchanged in the composite UR model \( \hat{y}^{(k)}_{UR} \) (i.e. \( \hat{\alpha}_{xi}^{(i)} = \lambda_{ext}^{(k)} \)).

5. UR models: Time-invariant confounder (Figure 2a)

We next consider the scenario in Figure 2a, in which a time-invariant covariate \( m \) confounds the relationship between \( x_1, x_2, \ldots, x_k \) and \( y \). This section is structured similarly to the preceding one. We provide: definitions of the standard regression models, UR terms, and UR models, all adjusted for the confounder \( m \) based upon the DAG in Figure 2a (§5.1); an analysis of UR models within a causal framework (§5.2); arguments for why Properties (i) – (iii) are upheld when the defined adjustments for \( m \) have been made (§5.3); and a discussion regarding the implications of insufficient adjustment for \( m \) (§5.4).

5.1. Definitions (with correct adjustment for \( m \))

Using the DAG in Figure 2a as guidance, we extend the original definitions of the standard regression models, UR terms, and UR models (Eq.5 – Eq.7, respectively) to properly account for the confounding effect of \( m \), a time-invariant covariate.

We define the OLS regression model \( \hat{y}^{(i)}_S \) for estimating the total causal effect of each measurement of the exposure variable \( x_i \) (for \( 1 \leq i \leq k \)) on \( y \) as:
\[ y_S^{(i)} = \hat{\alpha}_0^{(i)} + \hat{\alpha}_m^{(i)} m + \hat{\alpha}_x^{(i)} x_1 + \hat{\alpha}_x^{(i)} x_2 + \ldots + \hat{\alpha}_x^{(i)} x_i. \]  
(Eq.9)

Because the relationship between each \( x_i \) and \( y \) is confounded by all previous measurements of \( x \) (i.e. \( x_1, \ldots, x_{i-1} \)) and \( m \), these covariates must be adjusted for to obtain an inferentially unbiased estimate of the total causal effect of each measurement of the exposure. As previously, only the coefficient of the last/most recent measurement of \( x \) (i.e. \( \hat{\alpha}_x^{(i)} \)) may be interpreted as a total causal effect.

We further extend the process of Keijzer-Veen et al.\(^8\) to create UR terms for this scenario. As is evident, the relationship between each measurement of the exposure variable \( x_i \) and all previous measurements \( x_1, \ldots, x_{i-1} \) is confounded by \( m \) (for \( 2 \leq i \leq k \)); thus, adjustment for \( m \) is necessary:

\[ x_i = \hat{\phi}_0^{(i)} + \hat{\phi}_m^{(i)} m + \hat{\phi}_x^{(i)} x_1 + \hat{\phi}_x^{(i)} x_2 + \ldots + \hat{\phi}_x^{(i)} x_{i-1} + e_{xi}. \]  
(Eq.10)

Therefore, the UR term \( e_{xi} \) represents the difference between the actual value of \( x_i \) and the value of \( x_i \) as predicted by all previous measurements \( x_1, \ldots, x_{i-1} \), adjusted for the confounding effect of \( m \).

Finally, we define the UR model \( \hat{\gamma}_U^{(i)} \) (for \( 1 \leq i \leq k \)); this model must be also be adjusted for \( m \), since \( m \) confounds the relationship between \( x_i \) and \( y \):

\[ \hat{\gamma}_U^{(i)} = \hat{\lambda}_0^{(i)} + \hat{\lambda}_m^{(i)} m + \hat{\lambda}_x^{(i)} x_1 + \hat{\lambda}_x^{(i)} x_2 + \ldots + \hat{\lambda}_x^{(i)} x_k + e_{xi}. \]  
(Eq.11)

The composite UR model \( \hat{\gamma}_U^{(k)} \) thus represents the outcome \( y \) as function of the initial value of the exposure \( x_1 \), all subsequent ‘unexplained’ increases \( e_{x2}, \ldots, e_{xk} \), and the time-invariant confounder \( m \).

As in the preceding section, visual depictions of the previous equations are provided, with Figure 2b corresponding to Eq.8 and Figure 2c corresponding to Eq.8 and Eq.9 (with \( i = k \).
5.2. A causal framework

We may easily extend the reasoning from the previous scenario (§4.2) to explain why the UR model (Eq.11) satisfies Properties (i) – (iii) before resorting to mathematics, by considering the diagram in Figure 2a as a path diagram. A regression model containing all of \( m, x_1, x_2, \ldots, x_k \) (as in Eq.9) would only allow for the interpretation of the coefficient of \( x_k \) as a total causal effect on \( y \); the coefficients of \( x_1, \ldots, x_{k-1} \) would represent only the direct effects of each measurement on \( y \), because all future measurements would mediate the respective relationship and all backdoor paths would be blocked by preceding measurements (including \( m \)). Within the UR model, the independence of all UR terms \( e_{x2}, \ldots, e_{xk} \) ensures no mediating paths are blocked, and the only backdoor path between \( x_1 \) and \( y \) is blocked by \( m \).

5.3. Covariate orthogonality and Properties (i) – (iii)

In addition to the graph-based approach in the preceding section (§5.2), we are able to illustrate mathematically that adjustment for \( m \) both when generating each UR term \( e_{x1} \) (Eq.10) and in the composite UR model (Eq.11) will result in Properties (i) – (iii) being satisfied. Note that the scenario depicted in Figure 2a is nearly indistinguishable, both visually and mathematically, from the scenario in Figure 1a. The confounder \( m \) (which affects \( y \) and all measurements of \( x \)) could be reimagined as variable \( x_0 \); viewed in this way, the need for its adjustment becomes clear and the proofs from the previous section apply with only minor notational adjustments. Even though a distinction must be drawn between exposure variables and confounding variables within a causal framework, OLS regression treats both equivalently (i.e. as ‘independent variables’). Therefore, we give a brief outline
only of how the adjustments deemed necessary by the causal diagram in Figures 2a will result in Properties (i) – (iii) being upheld and attach the formal mathematical proofs in Appendix 2.

Eq.9 – Eq.11, which are summarised in Table 3, are guaranteed satisfy Properties (i) – (iii). As in the previous scenario (§4.3), each regression model (for both the standard and UR methods) in Table 3 contains one more covariate than the model preceding it – an additional $x_i$ term in the column of standard regression models, and an additional $e_{xi}$ term in the column of UR models. Proofs for the previous scenario relied on the property of each UR term being orthogonal to all preceding terms in the model. Adjustment for $m$ when generating each UR term $e_{xi}$ (Eq.10) guarantees that this property will be upheld, because it ensures that $e_{xi}$ is orthogonal to $m$ in addition to $e_{x1}, \ldots, e_{x(i-1)}$: this cannot be guaranteed without explicit adjustment for $m$. Furthermore, adjustment for $m$ in each UR model in Table 3 ensures that $\hat{\lambda}_{\text{UR}} = \hat{\lambda}_{\text{UR}}$ for each row in Table 3.

5.4. Incorrect adjustment for $m$

We have used the causal diagram in Figure 2a to argue for the necessity of adjusting for a time-invariant confounder $m$ during both stages of the UR modelling process, and have demonstrated how such adjustments will produce a composite UR model that satisfies Properties (i) – (iii), as Keijzer-Veen et al. intended. We now consider the implications of insufficient adjustment.
Without adjustment for \( m \) when generating each UR term \( e_{x_1} \), the coefficients of \( x_1, \ldots, x_{i-1} \) (i.e. \( \hat{y}_x^{(j)} \), for \( 1 \leq i \leq k - 1 \) and \( 1 \leq j \leq k \)) and the UR term will absorb the effect of the omitted variable \( m \) on \( x_i \), thereby biasing the total effect of \( e_{x_1} \) estimated within the UR model (so-called ‘omitted variable bias’). Further, it is evident that \( m \) confounds the relationship between \( x_1 \) and \( y \), so that failure to adjust for \( m \) in the composite UR model will produce different predicted outcomes and bias the estimated coefficient of \( x_1 \).

### 6. UR models: Time-varying confounder (Figure 3a)

Finally, we consider the scenario in Figure 3a, in which a time-varying covariate \( m_1, m_2, \ldots, m_k \) confounds the relationship between \( x_1, x_2, \ldots, x_k \) and \( y \).

In this section, we again provide: definitions of the standard regression models, UR terms, and UR models, all adjusted for the confounder \( m_1, m_2, \ldots, m_k \) based upon the DAG in Figure 3a (§6.1); an analysis of UR models within a causal framework (§6.2); arguments for why Properties (i) – (iii) are upheld when the defined adjustments for \( m_1, m_2, \ldots, m_k \) have been made (§6.3); and a discussion regarding the implications of insufficient adjustment for \( m_1, m_2, \ldots, m_k \) (§6.4).

#### 6.1. Definitions (with correct adjustment for \( m_1, m_2, \ldots, m_k \))

Using the DAG in Figure 3a, we extend the original definitions of the standard regression models, UR terms, and UR models (Eq.5 – Eq.7, respectively) to properly account for the confounding effect of \( m_1, m_2, \ldots, m_k \), a time-varying covariate.

We define the OLS regression model \( \hat{y}_x^{(i)} \) for estimating the total causal effect of each measurement of the exposure variable \( x_i \) (for \( 1 \leq i \leq k \)) on \( y \) as:
\[ y_S^{(i)} = \hat{\alpha}_0 + \hat{\alpha}_m^{(i)} m_1 + \hat{\alpha}_{x1}^{(i)} x_1 + \cdots + \hat{\alpha}_{m_i}^{(i)} m_i + \hat{\alpha}_{xi}^{(i)} x_i. \]  

(Eq.12)

The relationship between each \( x_i \) and \( y \) is not only confounded by all previous values of the exposure \( x_1, \ldots, x_{i-1} \) but also by the current measurement and all previous measurements of the confounder \( m_1, \ldots, m_i \). Therefore, adjustment for \( m_1, \ldots, m_i, x_1, \ldots, x_{i-1} \) is necessary to obtain an inferentially unbiased estimate of the total causal effect of each measurement of the exposure. We reiterate that only the coefficient of the last/most recent measurement of \( x \) (i.e. \( \hat{\alpha}_{xi}^{(i)} \)) may be interpreted as a total causal effect.

Extending the process of Keijzer-Veen et al.\(^9\) to create UR terms for each measurement of the exposure \( x_i \) in this scenario necessitates adjustment for the current measurement and all previous measurements of the confounder \( m_1, m_2, \ldots, m_i \) (for \( 2 \leq i \leq k \)), since these variables confound the relationship between each measurement of the exposure variable \( x_i \) and all previous measurements \( x_1, \ldots, x_{i-1} \), i.e.:

\[ x_i = \hat{\gamma}_0^{(i)} + \hat{\gamma}_{m1}^{(i)} m_1 + \hat{\gamma}_{x1}^{(i)} x_1 + \cdots + \hat{\gamma}_{m(i-1)}^{(i)} m_{i-1} + \hat{\gamma}_{xi}^{(i)} x_{i-1} + \hat{\gamma}_{m_i}^{(i)} m_i + e_{xi}. \]  

(Eq.13)

In this way, \( e_{xi} \) represents the difference between the observed value of \( x_i \) and the value of \( x_i \) as predicted by all previous measurements \( x_1, \ldots, x_{i-1} \), adjusted for the confounding effects of \( m_1, m_2, \ldots, m_i \).

As we have demonstrated previously (§4.3, §5.3), UR models rely upon the orthogonality of terms in the composite UR model. This necessitates the creation of UR terms \( e_{mi} \) for each measurement of the time-varying confounding variable \( m_i \) (for \( 2 \leq i \leq k \)) in a similar manner to that of the UR terms \( e_{xi} \) (Eq.13).

Each \( e_{mi} \) is derived from the OLS regression of \( m_i \) on all previous values of the confounder \( m_1, \ldots, m_{i-1} \), as well as all previous values of the exposure \( x_1, x_2, \ldots, x_{i-1} \) which confound this relationship:
\[ m_i = \hat{h}_0^{(i)} + \hat{h}_{m1}^{(i)} m_1 + \hat{h}_{x1}^{(i)} x_1 + \cdots + \hat{h}_{m(i-1)}^{(i)} m_{i-1} + \hat{h}_{x(i-1)}^{(i)} x_{i-1} + e_{mi}. \]  
(Eq.14)

Thus, \( e_{mi} \) has a similar interpretation to the original UR term \( e_{xii} \), in that it represents the part of \( m_i \) unexplained by all previous values \( m_1, \ldots, m_{i-1}, \) adjusted for the confounding effects of \( x_1, \ldots, x_{i-1} \).

Lastly, we define the UR model \( \hat{y}_{UR}^{(i)} \) (for \( 1 \leq i \leq k \)) as a function of the initial value of the confounder \( m_1 \) and its subsequent ‘unexplained’ increases \( e_{m2}, \ldots, e_{mi} \), and the initial value of the exposure \( x_1 \) and its subsequent ‘unexplained’ increases \( e_{x2}, \ldots, e_{xi} \):

\[ \hat{y}_{UR}^{(i)} = \hat{\lambda}_0^{(i)} + \hat{\lambda}_{m1}^{(i)} m_1 + \hat{\lambda}_{x1}^{(i)} x_1 + \hat{\lambda}_{em2}^{(i)} e_{m2} + \hat{\lambda}_{ext}^{(i)} e_{x2} + \cdots + \hat{\lambda}_{eml}^{(i)} e_{mi} + \hat{\lambda}_{ext}^{(i)} e_{xi}. \]  
(Eq.15)

As previously, visual depictions of these equations are provided. Figure 3b corresponds to the standard regression models given by Eq.12; Figure 3c corresponds to the \( k - 1 \) regressions of \( x_i \) on all preceding measurements of \( x \) and \( m \) (Eq.13), the \( k - 1 \) regressions of \( m_i \) on all preceding measurements of \( x \) and \( m \) (Eq.14), and one composite UR regression model (Eq.15, with \( i = k \)).

6.2. A causal framework

The similarities amongst the three causal scenarios depicted in Figures 1a, 2a, and 3a are evident, and shed light on how the reasoning from the previous scenarios (§4.2 and §5.2) can be extended to demonstrate why the UR model in Eq.15 satisfies Properties (i) – (iii). In a regression model containing all of \( m_1, \ldots, m_k, x_1, \ldots, x_k \) (as in Eq.12, with \( i = k \)), only the coefficient of \( x_k \) could be interpreted as a total causal effect on \( y \); the coefficients of \( x_1, \ldots, x_{k-1} \) may only be interpreted as the direct effects of each measurement of the exposure on \( y \), because all future measurements of both \( x \) and \( m \) would fully mediate the respective relationship and all preceding measurements of \( x \) and \( m \) would block all backdoor paths. Within the UR model, however, the independence of all UR terms for both the
exposure (i.e. $e_{x_2}, \ldots, e_{x_k}$) and mediator (i.e. $e_{m_2}, \ldots, e_{m_k}$) ensures no mediating paths are blocked, and the only backdoor path between $x_1$ and $y$ is blocked by $m_4$.

6.3. Covariate orthogonality and Properties (i) – (iii)

In addition to the graph-based approach in the preceding section (§6.2), we can illustrate mathematically that the standard regression models $\hat{\gamma}_S^{(i)}$ (Eq.12), UR terms for measurements of the exposure (Eq.13) and confounder (Eq.14), and composite UR model $\hat{\gamma}_{UR}^{(k)}$ (Eq.15, with $i = k$) satisfy Properties (i) – (iii). Although seemingly more complex, the scenario depicted in Figure 3a also has very little to distinguish it from the scenarios in Figures 1a and 2a. The confounder $m_1$, being the only exogenous node on the graph, could be imagined as variable $x_0$, with all nodes subsequent to $x_1$ having an associated UR term. Viewed as such, the necessity of adjusting for $m_1$ and creating UR terms for both the exposure and the time-varying confounder becomes apparent, as the causal diagram in Figure 3a is equivalent to that of Figure 2a with minor notational adjustments. Therefore, we provide only a brief outline of how the adjustments deemed necessary by the causal diagrams in Figures 3a will result in Properties (i) – (iii) being upheld; formal mathematical proofs are provided in Appendix 3.

Eq.12 – Eq.15 are summarised in Table 4 and are guaranteed to satisfy Properties (i) – (iii). In contrast to previous scenarios (§4.3 and §5.3), each regression model (for both the standard and UR models) contains two more covariates than the model preceding it. In the column of standard regression models, each row contains an additional $x_i$ and $m_i$ term; in the column of UR models, each row contains an additional $e_{x_i}$ and $e_{m_i}$ term. Thus, for Properties (i) – (iii) to be upheld in each UR model $\hat{\gamma}_{UR}^{(i)}$, these two additional terms must be orthogonal to one another and to all preceding terms.
\begin{align*}
\text{i} = 1: & \quad \hat{\alpha}_0^{(1)} + \hat{\alpha}_{m1}^{(1)} m_1 + \hat{\alpha}_{x1}^{(1)} x_1 \\ & \quad \hat{\lambda}_0^{(1)} + \hat{\lambda}_{m1}^{(1)} m_1 + \hat{\lambda}_{x1}^{(1)} x_1 \\
\text{i} = 2: & \quad \hat{\alpha}_0^{(2)} + \hat{\alpha}_{m1}^{(2)} m_1 + \hat{\alpha}_{x1}^{(2)} x_1 + \hat{\alpha}_{m2}^{(2)} m_2 + \hat{\alpha}_{x2}^{(2)} x_2 \\ & \quad \hat{\lambda}_0^{(2)} + \hat{\lambda}_{m1}^{(2)} m_1 + \hat{\lambda}_{x1}^{(2)} x_1 + \hat{\alpha}_{m2}^{(2)} m_2 + \hat{\lambda}_{x2}^{(2)} x_2 \\
\vdots & \quad \vdots \\
\text{i} = k: & \quad \hat{\alpha}_0^{(k)} + \hat{\alpha}_{m1}^{(k)} m_1 + \hat{\alpha}_{x1}^{(k)} x_1 + \hat{\alpha}_{m2}^{(k)} m_2 + \hat{\alpha}_{x2}^{(k)} x_2 + \cdots + \hat{\alpha}_{m_{k-1}}^{(k)} m_{k-1} + \hat{\alpha}_{x_{k-1}}^{(k)} x_{k-1} \\ & \quad \hat{\lambda}_0^{(k)} + \hat{\lambda}_{m1}^{(k)} m_1 + \hat{\lambda}_{x1}^{(k)} x_1 + \hat{\alpha}_{m2}^{(k)} m_2 + \hat{\lambda}_{x2}^{(k)} x_2 + \cdots + \hat{\alpha}_{m_{k-1}}^{(k)} m_{k-1} + \hat{\lambda}_{x_{k-1}}^{(k)} x_{k-1} 
\end{align*}

Table 4: For the scenario depicted in Figures 3a–b, the standard regression model \( \hat{\gamma}_S^{(i)} \) necessary for estimating the total causal effect of each exposure \( x_i \) on \( y \), and the corresponding UR model \( \hat{\gamma}_{UR}^{(i)} \) for \( 1 \leq i \leq k \).

Proving this is relatively straightforward. For any UR term \( e_{mi} \) for the confounder, it holds that \( e_{mi} \) is orthogonal to \( m_1, \ldots, m_{i-1}, x_1, \ldots, x_{i-1} \) by construction (Eq.14); because preceding UR terms \( e_{x2}, \ldots, e_{x(i-1)} \) (Eq.13) and \( e_{m2}, \ldots, e_{m(i-1)} \) (Eq.14) may be expressed as linear combinations of \( m_1, \ldots, m_i, x_1, \ldots, x_{i-1} \), it follows that \( e_{mi} \) is orthogonal to \( e_{m2}, \ldots, e_{m(i-1)}, e_{x2}, \ldots, e_{x(i-1)} \). Furthermore, for any UR term \( e_{xi} \) for the exposure, it holds that \( e_{xi} \) is orthogonal to \( m_1, \ldots, m_{i-1}, x_1, \ldots, x_{i-1} \) by construction (Eq.13); because preceding UR terms \( e_{x2}, \ldots, e_{x(i-1)} \) (Eq.13) and \( e_{m2}, \ldots, e_{m(i-1)} \) (Eq.14) may be expressed as linear combinations of \( m_1, \ldots, m_i, x_1, \ldots, x_{i-1} \), it follows that \( e_{xi} \) is orthogonal to \( e_{m2}, \ldots, e_{m(i-1)}, e_{x2}, \ldots, e_{x(i-1)} \). Thus, we are able to conclude that \( e_{mi} \) and \( e_{xi} \) are orthogonal to each other and to all preceding terms in for any UR model \( \hat{\gamma}_{UR}^{(i)} \), adjustment for all causally preceding measurements of both \( m \) and \( x \) when generating UR terms for both the confounder and the exposure ensures this orthogonality.

6.4. Incorrect adjustment for \( m_1, m_2, \ldots, m_k \)

The DAG in Figure 3a demonstrates the necessity of adjusting for a time-varying confounder \( m_1, m_2, \ldots, m_k \) in the manner described in Section 6.1, and we have demonstrated how such adjustments will produce a composite UR model that satisfies Properties (i) – (iii). The implications of incorrect adjustment for a time-varying confounder \( m_1, m_2, \ldots, m_k \) in a UR model are similar to those of
incorrect adjustment for a time-invariant confounder \( m \), which were previously outlined in Section 5.4.

Without adjustment for any of \( m_1, \ldots, m_i \) when constructing each UR term for the exposure \( e_{xi} \), the coefficients of \( x_1, \ldots, x_{i-1} \) (i.e. \( \gamma^{(j)}_{xi} \), for \( 1 \leq i \leq (k - 1) \) and \( 1 \leq j \leq k \)) and the UR term will absorb the effect of each omitted variable on \( x_i \); this will result in the coefficient estimated for each \( e_{xi} \) in the composite UR model to be unequal to the total effect of \( x_i \) in its corresponding standard regression model.

The requirement of orthogonal covariates within the composite UR model also sheds light on the necessity for generating UR terms \( e_{m2}, e_{m3}, \ldots, e_{mk} \) for measurements of a time-varying confounder, if present. We might easily imagine a scenario in which we considered only the original covariates \( m_1, m_2, \ldots, m_k \) in the UR model. In such a scenario, the terms would remain correlated with each other and with \( x_1 \); therefore, the inclusion of subsequent \( m \) terms in the UR model would necessarily change the coefficient estimates for \( x_1 \) and all other covariates.

7. UR model interpretation

Having demonstrated that confounder adjustment within UR models is possible, we consider the claim\(^9\) that UR models offer additional insight (via the coefficients for each UR term \( e_{xi} \) (i.e. \( \hat{y}_{ext}^{(k)} \) in Eq.7, for \( 2 \leq i \leq k \)) into the effect of \( x_i \) increasing more than expected upon \( y \).

Consider again the simple example with two longitudinal measurements of a continuous exposure \( x \) (i.e. \( x_1 \) and \( x_2 \)), outcome \( y \), and no additional confounders (i.e. Figure 1a, with \( k = 2 \)); the standard regression model (with \( x_2 \) as the specified exposure variable) and ‘equivalent’ UR model are given below, respectively:
\[ \hat{y}^{(2)}_S = \hat{a}^{(2)}_0 + \hat{a}^{(2)}_{x_1} x_1 + \hat{a}^{(2)}_{x_2} x_2 \]

\[ \hat{y}^{(2)}_{UR} = \hat{a}^{(2)}_0 + \hat{a}^{(2)}_{x_1} x_1 + \hat{a}^{(2)}_{exx2} e x_2 . \]

It has been shown (§4.3) that \( \hat{a}^{(2)}_{x_2} \) and \( \hat{a}^{(2)}_{exx2} \) are equal, yet \( \hat{a}^{(2)}_{x_2} \) is interpreted as the total effect of a one-unit increase in \( x_2 \) on \( y \), whereas \( \hat{a}^{(2)}_{exx2} \) (supposedly) interpreted as the total effect of a one-unit higher than expected increase in \( x_2 \) on \( y \). If these two variables truly are distinct, their regression coefficients should likewise be distinct. This issue has also been addressed by Tu and Gilthorpe, who have argued that the two coefficients are equivalent because adjustment for \( x_1 \) in \( \hat{y}^{(2)}_S \) amounts to testing the relation between \( y \) and the part of \( x_2 \) unexplained by \( x_1 \) (i.e. the unexplained residual). In fact, the two coefficients are equal simply because they mean the same thing. The UR model does not, therefore, offer any additional insight into the effect of higher than expected change in \( x \) on the outcome.

We also raise a more philosophical point, which speaks to the need for any model to reflect accurately the underlying data-generation process of a given scenario. As an artefact of OLS regression, the UR terms will always be mathematically independent of the value of the initial measurement of the exposure and all subsequent measurements. This is unlikely to be an accurate representation of real-world exposure variables. Many of these, such as body size, exhibit a consistent, cumulative presence that is only manifest at the discrete time points at which it is measured; these measurements are thus distinct only as a result of the discretisation of time within the measurement processes adopted.

Moreover, in auxological studies, the phenomenon of so-called compensatory (or ‘catch up’) growth has been well documented, with accelerated growth being observed in individuals who begin with a low value of some measure, e.g. birthweight. Therefore, however convenient and mathematically sound it may be to model data in a way that implies complete statistical independence amongst an
exposure variable’s initial value and its subsequent measurements, this assumption is likely to be implausible and unrealistic for most biological and social variables of interest to epidemiologists. This is a weakness shared by all conditional approaches (of which UR models are one), which has led several authors\textsuperscript{47} to recommend that the results be considered alongside those produced by other methods, rather than in isolation.

8. Standard error reduction

Finally, we address an important consequence of the use of UR models; namely, that they underestimate the standard errors (SEs) of estimated coefficients, thereby resulting in artificial precision of estimated effect sizes. Although focus on statistical significance by way of p-values and confidence intervals is not in and of itself justifiable within a causal framework (as focus is effect size and likely functional significance, e.g. the absolute risk posed or the potential for substantive intervention), we consider it an important issue to address as a matter of clarity for researchers seeking to use UR models.

To demonstrate, we have simulated 1000 non-overlapping random samples of 1000 observations from a multivariate normal distribution based upon the DAG in Figure 1a with $k = 2$, using the ‘dagitty’ package (v. 0.2-2)\textsuperscript{4, 48} in R (v. 3.3.2)\textsuperscript{49}. Each sample was used to create: (1) the two standard regression models necessary for estimating the total causal effect of each of $x_1, x_2$ on $y$ (Eq.5); (2) the UR term $e_{x_2}$, derived by regressing $x_2$ on $x_1$ (Eq.6); and (3) the composite UR model in which $y$ is regressed on $x_1$ and $e_{x_2}$ (Eq.7). For each standard regression model $\hat{\beta}^{(i)}_S$ (for $i = 1,2$), the reported SE of the regression coefficient for exposure $x_i$ is stored. For each composite UR model $\hat{\beta}^{(2)}_UR$, the SE of the regression coefficient for each of $x_1, e_{x_2}$ is stored in two forms: (1) as reported in the UR model summary output; and (2) as estimated by bootstrapping 1000 samples and calculating the standard deviation of the
distribution of estimated coefficients. Additional details relating to this simulation – including parameters and code – are located in Appendix 4.

By definition, the SE of an estimated regression coefficient is a point estimate of the standard deviation of an (infinitely) large sampling distribution of estimated regression coefficients. We have shown that standard regression and UR models elicit identical point estimates of the total causal effects of each measure of the longitudinal exposure (§4); from this, it follows that the associated SEs should themselves be equal.

[Insert Figure 4a-b]

Violin plots of the SEs estimated for each coefficient representing a total causal effect across the 1000 simulations are displayed in Figure 4 for each method considered. As is evident, the reported SEs within the UR models are reduced in comparison to those within the first standard regression models (for designated exposure $x_1$) and equal to those within the final standard regression models (for designated exposure $x_2$). This demonstrates an apparent paradox: the coefficient values are equivalent, yet the associated SEs are unequal.

We argue that the apparent reduction in standard errors achieved by using UR models is purely artefactual and arises from the explicit conditioning on future measurements of $x$ within a UR model. In the standard regression analysis, the only information within the data that is used to inform SE estimation lies in the past (i.e. past measures of the exposure plus any confounders). In contrast, the UR modelling process generates (orthogonal) residuals for the entire exposure period and combines these into a single model, thereby using information within the data that is from both the past and the future.

---

$^b$ The specific correlation structure and parameter values used to simulate the data are unimportant for the purposes of this demonstration.
If we possessed data pertaining to any true independent causes of future measurements of the exposure, such a method would indeed be valid; however, the UR terms are simply estimated using prior measurements of the exposure. Moreover, due to the fact that they are estimates, the UR terms themselves contain additional variation that is not accommodated by traditional regression methods which assume covariates are measured without error. Consequently, the SEs of estimated causal effect derived from UR models are artefactually reduced and should not be inferred as robust. Indeed, when the SEs within the UR models are estimated via bootstrapping, they are similar to those within the standard regression models.

Comparing the two plots in Figure 4 offers clarity to this argument: (a) displays differing distributions of the reported SEs for the coefficient estimates of $x_1$ (where conditioning on the future information given by $x_2$ reduces the standard error in the UR model); whereas (b) displays the same distribution of the reported SEs for the coefficient estimates of $x_2$ and $e_{x_2}$ (where the standard regression model correctly exploits all prior information given by $x_1$, as does the UR model). Although the magnitude of bias in estimated SEs is small in this simulated example, it will always be present due to the way in which UR models are constructed. Quantifying the magnitude of this bias is not trivial and is beyond the scope of the present study.

9. Conclusion

The mathematical appraisal of UR models that we have undertaken confirms that the method proposed by Keijzer-Veen et al. is capable of accommodating more than two longitudinal measurements of an exposure variable and demonstrates how adjustment for confounding variables should be made in this framework to uphold the property that the coefficients for the terms $x_1, e_{x_2}, \ldots, e_{x_k}$ estimated within a
UR model are equal to the total effects for $x_1, x_2, \ldots, x_k$ estimated by their respective standard
regression models. This result will only be guaranteed to hold when adjustment for all confounding
variables has been made at both stages in the UR modelling process (i.e. when generating UR terms for
subsequent measurements of the exposure and in the composite UR model). From a statistical
perspective, adjustment for all preceding variables (including confounders) ensures orthogonality
amongst the covariates in a composite UR model. Therefore, when the potential confounder is time-
varying, it is also necessary to generate UR terms for subsequent measurements of the confounder itself
and include these in the final composite models used.

As our proofs only consider one confounding variable, the causal framework provided by DAGs should
aid future researchers who wish to extend robustly UR models to situations involving multiple, possibly
causally linked, time-invariant and time-varying confounders. Such a DAG will be useful in identifying
confounders and establishing the temporal ordering of variables, thereby ensuring that all preceding
variables are adjusted for when generating the necessary UR terms.

Although UR models can accommodate multiple measurements of an exposure variable in addition to
confounding variables, we have concerns about their practical implementation. Although only one UR
model need ultimately be presented, the necessity of generating orthogonal covariates for that UR
model requires that many models be created; this has the potential to be quite substantial when
multiple confounders are considered. For an exposure $x$ measured at $k$ points in time, the standard
regression approach necessitates $k$ separate models for estimating the total causal effect of each
measurement on the outcome regardless of the number of confounders. In the case of one time-
invariant confounder ($\S 5$), $k$ models are also created ($k - 1$ models to generate all UR terms and 1
composite UR model); for a time-varying confounder ($\S 6$), $2k - 1$ models are created (i.e. $2k - 2$
models to generate all UR terms and 1 composite UR model). The total number of models created by
the UR process will always be either equal to or greater than the total number of models created by the
standard regression process. If such a process offered real gains in insight into the scenario under
consideration, it may indeed be worth it; however, UR models offer no additional insight compared to
standard regression methods. Moreover, the inclusion of multiple covariates that are explicitly
conditional on one another within the same model also results in artificially reduced standard error
estimates, the extent of which has yet to be fully evaluated; the issue can be avoided by bootstrapping,
but such a solution may be computationally intensive and require more programming skills than those
necessary for implementing the built-in regression functionalities in statistical software packages.
Previous research that has utilised UR models without undertaking sufficient adjustment for
confounders and correcting standard errors via bootstrapping should not be considered robust.

We therefore have strong reservations about the use and implementation of UR models within
lifecourse epidemiology, and suggest that researchers considering using them should instead rely on
standard regression methods, which produce the same results but are much less likely to be mis-
specified and misleading. However, for researchers wishing to use these models, the hypothesised DAG
or causal diagram should be presented so that any readers and/or reviewers can confirm that sufficient
adjustment for confounders has been undertaken; moreover, standard errors should be estimated via
bootstrapping and not simply reported as in the model output, as these have the potential to be
misleading. We support the recommendation of previous authors that additional analytical
approaches should be considered alongside conditional approaches (e.g. UR models) in order to achieve
robust causal conclusions. For example, multilevel, latent growth curve, and growth mixture models
may be used to estimate the effects of growth across the lifecourse on a distal outcome, and are more
flexible than standard regression methods. Moreover, the three G-methods are explicitly grounded
in a causal framework and allow for the simultaneous consideration of multiple measurements of a longitudinally measured exposure, as well as time-varying confounding; these methods provide exciting avenues of research for lifecourse epidemiologists.
**Funding**

KFA and SCG were supported by the Economic and Social Research Council [grant numbers ES/J500215/1 and ES/P000746/1, respectively]. GTHE, PWGT, AH, and MSG were supported by the Higher Education Funding Council for England.

**Conflict of interest**

The Authors declare that there is no conflict of interest.

**References**

7. Tu YK, Gilthorpe MS and Ellison GTH. What is the effect of adjusting for more than one measure of current body size on the relation between birthweight and blood pressure? *Journal of Human Hypertension*. 2006; 20: 646-57.


1. Appendix 1: UR models: No confounders

Proofs of Properties (i) – (iii) for the scenario depicted in Figure 1a (i.e. k longitudinally measured exposure variables $x_1, x_2, ..., x_k$ and one distal outcome $y$).

1.1. Definitions

1.1.1. Definition 1: Standard regression models

We define the ordinary least-squares (OLS) regression model $\hat{\beta}_S^{(i)}$ for each measurement of the exposure variable $x_i$, for $1 \leq i \leq k$. Because the relationship between $x_i$ and $y$ is confounded by all previous values of $x$ (i.e. $x_1, x_2, ..., x_{i-1}$), we represent $y$ as a function of $1, x_1, x_2, ..., x_i$:

$$\hat{\beta}_S^{(i)} = \hat{\alpha}_0 + \hat{\alpha}_{x_1} x_1 + \hat{\alpha}_{x_2} x_2 + \cdots + \hat{\alpha}_{x_k} x_k.$$  

(Eq.1)

As discussed in Section 1, only the coefficient of the last/most recent measurement of $x$ (i.e. $\hat{\alpha}_x$) may be interpreted as a total causal effect.

1.1.2. Definition 2: Unexplained residual (UR) terms

As established by Keijzer-Veen et al., each UR term $e_{xi} is derived from the OLS regression of $x_i$ on all previous measurements of $x$ (i.e. $x_1, x_2, ..., x_{i-1}$):

$$x_i = \hat{\beta}_0^{(i)} + \hat{\beta}_{x_1}^{(i)} x_1 + \hat{\beta}_{x_2}^{(i)} x_2 + \cdots + \hat{\beta}_{x_{i-1}}^{(i)} x_{i-1} + e_{xi}.$$  

(Eq.2)

for $2 \leq i \leq k$. Thus,

$$e_{x2} = -\hat{\beta}_0^{(2)} - \hat{\beta}_{x1}^{(2)} x_1 + x_2$$

$$e_{x3} = -\hat{\beta}_0^{(3)} - \hat{\beta}_{x1}^{(3)} x_1 - \hat{\beta}_{x2}^{(3)} x_2 + x_3$$

$$\vdots$$

$$e_{xk} = -\hat{\beta}_0^{(k)} - \hat{\beta}_{x1}^{(k)} x_1 - \hat{\beta}_{x2}^{(k)} x_2 - \cdots - \hat{\beta}_{x_{k-1}}^{(k)} x_{k-1} + x_k.$$  

(Eq.3)

By its formulation, $e_{xi}$ represents the difference between the actual value of $x_i$ and the value of $x_i$ as predicted by all previous measurements of $x$.

1.1.3. Definition 3: Unexplained residuals (UR) models

We also define the UR model $\hat{\beta}_U^{(i)}$ – an OLS regression model which represents $y$ as a function of $1, x_1, e_{x2}, ..., e_{xi}$, for $1 \leq i \leq k$ – as:

$$\hat{\beta}_U^{(1)} = \hat{\beta}_0^{(1)} + \hat{\beta}_{x1}^{(1)} x_1$$

$$\hat{\beta}_U^{(2)} = \hat{\beta}_0^{(2)} + \hat{\beta}_{x1}^{(2)} x_1 + \hat{\beta}_{ex2}^{(2)} e_{x2}$$

https://mc.manuscriptcentral.com/smr
\[ y^{(k)}_{UR} = \hat{y}^{(k)}_0 + \hat{\lambda}^{(k)}_{x_1} x_1 + \hat{\lambda}^{(k)}_{x_2} e_{x_2} + \cdots + \hat{\lambda}^{(k)}_{e_{x_k}} e_{x_k}. \]  

(Eq.4)

Thus, the final composite model \( y^{(k)}_{UR} \) represents the outcome as a function of the initial value of \( x \) and all subsequent increases.

1.2. Mathematical proofs

The proofs that follow rely upon the following key properties of OLS regression estimators and require the following two lemmas:

**Key properties of OLS estimators:** We may represent the regression equation \( y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon \) in summary notation as:

\[ y = X\beta + \epsilon, \]

where: \( y \) represents the vector of \( n \) continuous observations of the outcome; \( X \) represents the \( n \times (k + 1) \) matrix of \( n \) observations for \( k \) continuous covariates and 1 constant; \( \beta \) represents the \( k + 1 \) vector of coefficients for each covariate and constant; and \( \epsilon \) represents the vector of \( n \) residuals.

The OLS estimate of \( \beta \) is given by:

\[ \hat{\beta} = (X'X)^{-1}X'y. \]

On the assumption that the inverse matrix exists, this equation has a unique solution.

Further, for the given OLS equation \( y = X\hat{\beta} + \epsilon \), it can be shown that the vector of residuals \( (\epsilon) \) is orthogonal (denoted \( \perp \) ) to every column \( (1, x_1, x_2, \ldots, x_k) \) of \( X \).

*Note that detailed proofs have not been provided, but can be located in referenced material.*

**Lemma 1:** For two orthogonal components \( \tau \) and \( \delta \) (i.e. \( \tau \perp \delta \)), the estimated coefficients of the regression of \( y \) on \( \tau \) and \( \delta \) are equal to the estimated coefficients for the separate regressions of \( y \) on \( \tau \) and \( y \) on \( \delta \).

**Proof of Lemma 1:** The regression of \( y \) on \( \tau \) and \( \delta \) may be written as:

\[ y = [\tau \quad \delta] \begin{bmatrix} \beta_\tau \\ \beta_\delta \end{bmatrix} + \epsilon = \tau \beta_\tau + \delta \beta_\delta + \epsilon. \]

From Definition 1, the OLS estimate of \( \beta_\tau \) and \( \beta_\delta \) is given by \( \hat{\beta} = (X'X)^{-1}X'y \). In this scenario,

\[ X'X = \begin{bmatrix} \tau' \delta' \\ \delta' \delta \end{bmatrix} \begin{bmatrix} \tau \\ \delta \end{bmatrix} = \begin{bmatrix} \tau' \tau & \tau' \delta \\ \delta' \tau & \delta' \delta \end{bmatrix} = \begin{bmatrix} \tau' \tau & 0 \\ 0 & \delta' \delta \end{bmatrix}, \]

where the final equivalency follows from the condition of orthogonality. Then

\[ (X'X)^{-1} = \begin{bmatrix} \tau' \tau & 0 \\ 0 & \delta' \delta \end{bmatrix}^{-1} = \begin{bmatrix} (\tau' \tau)^{-1} & 0 \\ 0 & (\delta' \delta)^{-1} \end{bmatrix}, \]

and

\[ X'y = \begin{bmatrix} \tau' \\ \delta' \end{bmatrix} y = \begin{bmatrix} \tau'y \\ \delta'y \end{bmatrix}. \]

Combining these elements gives:
\[
\begin{bmatrix}
\hat{\beta}_\tau \\
\hat{\beta}_\delta \\
\end{bmatrix}
= \begin{bmatrix}
(t'\tau)^{-1} & 0 \\
0 & (\delta'\delta)^{-1}
\end{bmatrix}
\begin{bmatrix}
\tau'y \\
\delta'y
\end{bmatrix} = \begin{bmatrix}
(t'\tau)^{-1}t'y \\
(\delta'\delta)^{-1}\delta'y
\end{bmatrix}.
\]

From this, we see that the estimated coefficients are equivalent to those that would be produced for the separate regressions of \(y\) on \(\tau\) and \(y\) on \(\delta\). ■

**Lemma 2:** If \(\tau_i \perp \delta_j\) for \(0 \leq i \leq h\) and \(0 \leq j \leq k\), then \(span(\tau_0, \tau_1, ..., \tau_h) \perp span(\delta_0, \delta_1, ..., \delta_k)\) for any vectors \(\tau_0, \tau_1, ..., \tau_h, \delta_0, \delta_1, ..., \delta_k\).

**Proof of Lemma 2:** \(\tau_i \perp \delta_j\) implies that \(\tau_i \cdot \delta_j = 0\) for \(0 \leq i \leq h\) and \(0 \leq j \leq k\). Then

\[
span(\tau_0, \tau_1, ..., \tau_h) \cdot span(\delta_0, \delta_1, ..., \delta_k)
= (c_0\tau_0 + c_1\tau_1 + \cdots + c_h\tau_h) \cdot (d_0\delta_0 + d_1\delta_1 + \cdots + d_k\delta_k)
= c_0d_0(\tau_0 \cdot \delta_0) + c_0d_1(\tau_1 \cdot \delta_1) + \cdots + c_0d_k(\tau_0 \cdot \delta_k)
+ c_1d_1(\tau_1 \cdot \delta_1) + \cdots + c_1d_k(\tau_1 \cdot \delta_k)
+ \cdots + c_hd_0(\tau_h \cdot \delta_0) + c_hd_1(\tau_h \cdot \delta_1) + \cdots + c_hd_k(\tau_h \cdot \delta_k)
= c_0d_0(0) + c_0d_1(0) + \cdots + c_0d_k(0) + c_1d_0(0) + c_1d_1(0) + \cdots + c_1d_k(0)
+ \cdots + c_hd_0(0) + c_hd_1(0) + \cdots + c_hd_k(0)
= 0.
\]

Thus, \(span(\tau_0, \tau_1, ..., \tau_h) \perp span(\delta_0, \delta_1, ..., \delta_k)\). ■

### 1.2.1. Covariate orthogonality

We prove that all UR terms \(e_{x_2}, e_{x_3}, ..., e_{x_k}\) are orthogonal to all preceding variables in the composite UR model (Eq.3), and therefore orthogonal to their span; we prove this below.

**Lemma 3:** \(e_{x_1} \perp \{e_{x_2}, e_{x_3}, ..., e_{x_{i-1}}\}\) for \(2 \leq i \leq k\).

**Proof of Lemma 3:** By construction, \(e_i\) represents the residuals from the OLS regression of \(x_i \sim 1, x_1, x_2, ..., x_{i-1}\). Thus, \(e_{x_1} \perp 1, x_1, x_2, ..., x_{i-1}\), which implies that \(e_{x_1} \perp span(1, x_1, x_2, ..., x_{i-1})\) by Lemma 2.

It is clear that \(e_{x_2}, e_{x_3}, ..., e_{x_{i-1}} \in span(1, x_1, x_2, ..., x_{i-1})\) for \(2 \leq i \leq k\) by construction; we are therefore able to conclude that \(e_{x_1} \perp e_{x_2}, e_{x_3}, ..., e_{x_{i-1}}\). ■

**Theorem 1:** \(e_{x_1} \perp \{e_{x_2}, e_{x_3}, ..., e_{x_{i-1}}\}\) for \(2 \leq i \leq k\).

**Proof of Theorem 1:** \(e_{x_1} \perp 1, x_1\) because \(e_{x_1}\) represents the residuals from the OLS regression of \(x_i \sim 1, x_1, x_2, ..., x_{i-1}\). Further, \(e_{x_1} \perp e_{x_2}, e_{x_3}, ..., e_{x_{i-1}}\) for \(2 \leq i \leq k\) by Lemma 3.

Thus, \(e_{x_1} \perp \{e_{x_2}, e_{x_3}, ..., e_{x_{i-1}}\}\) by Lemma 2. ■

---

1 The span of a set of vectors \(\delta_0, \delta_1, \delta_2, ..., \delta_k\) is the set of all possible linear combinations of \(\delta_0, \delta_1, \delta_2, ..., \delta_k\), i.e.: 

\[
span(\delta_0, \delta_1, \delta_2, ..., \delta_k) = c_0\delta_0 + c_1\delta_1 + c_2\delta_2 + \cdots + c_k\delta_k,
\]

where the coefficients \(c_0, c_1, c_2, ..., c_k\) are scalars.
1.2.2. Property (i): \( y_{sk}^{(k)} = \hat{y}_{UR}^{(k)} \)

Proof of Property (i): This equality follows from the fact that each UR model \( y_{UR}^{(i)} \) is a function of the same variables as the corresponding standard regression model \( y_{sk}^{(i)} \).

By Definition 3, \( y_{UR}^{(i)} = f(1, x_1, e_{x2}, ..., e_{xi}) \), where \( e_{xi} = f(1, x_1, x_2, ..., x_i) \) by Definition 2. Thus, it also holds that

\[
\hat{y}_{UR}^{(i)} = f(1, x_1, x_2, ..., x_i).
\]

Moreover, by Definition 1,

\[
\hat{y}_{sk}^{(i)} = f(1, x_1, x_2, ..., x_i).
\]

From this, it follows that \( y_{sk}^{(i)} = \hat{y}_{UR}^{(i)} \) and, consequently, \( y_{sk}^{(k)} = \hat{y}_{UR}^{(k)} \). ■

1.2.3. Property (ii): \( \hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(k)} \)

Proof of Property (ii): By definition, \( y_{sk}^{(i)} = y_{UR}^{(i)} = f(1, x_1) \), and so it is trivially true that \( \hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(k)} \).

Because \( e_{xi} \perp \text{span}(1, x_1, e_{x2}, e_{x3}, ..., e_{x(i-1)}) \) for \( 2 \leq i \leq k \) by Theorem 1, we are able to apply Lemma 1 and conclude that \( \hat{\lambda}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(2)} = ... = \hat{\lambda}_{x1}^{(k)} \).

Therefore, \( \hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(k)} \). ■

1.2.4. Property (iii): \( \hat{\alpha}_{xt}^{(i)} = \hat{\lambda}_{xt}^{(k)} \)

Proof of Property (iii): Consider the UR model:

\[
\hat{y}_{UR}^{(i)} = \hat{\lambda}_{0}^{(i)} + \hat{\lambda}_{x1}^{(i)} x_1 + \hat{\lambda}_{x2}^{(i)} e_{x2} + ... + \hat{\lambda}_{xt}^{(i)} e_{xt}.
\]

If we substitute the expansion for \( e_{xt} \) (Eq.3) into this equation and rearrange, we produce:

\[
\hat{y}_{UR}^{(i)} = \hat{\lambda}_{0}^{(i)} + \hat{\lambda}_{x1}^{(i)} x_1 + \hat{\lambda}_{x2}^{(i)} e_{x2} - \hat{y}_{x1}^{(2)} x_1 + x_2 + ... + \hat{\lambda}_{xt}^{(i)} [e_{xt}^{(i)} x_1 - e_{xt}^{(i)} x_2] - ... - \hat{\lambda}_{x(t-i)}^{(i)} x_{t-i} + x_i.
\]

Since we have already established that \( y_{sk}^{(i)} = \hat{y}_{UR}^{(i)} \) (i.e. Property (i)) because they are functions of the same covariates, it follows that the estimated coefficients for those covariates must themselves be equal. Specifically, we are able to see that the coefficient for \( x_i \) will always equal the coefficient for \( e_{xi} \), i.e. \( \hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{xt}^{(k)} \).

Finally, because \( e_{xi} \perp \text{span}(1, x_1, e_{x2}, e_{x3}, ..., e_{x(i-1)}) \), we can again apply Lemma 1 and conclude that \( \hat{\lambda}_{xt}^{(1)} = \hat{\lambda}_{xt}^{(2)} = ... = \hat{\lambda}_{xt}^{(k)} \), from which it follows that \( \hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{xt}^{(k)} \). ■

---

2 Although no causal meaning/significance can be attributed to the intercept term, the logic applied in this proof may be easily extended to show that \( \hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{xt}^{(k)} \).
2. Appendix 2: UR models: Time-invariant confounder

Proofs of Properties (i) – (iii) for the scenario depicted in Figure 2a (i.e. \( k \) longitudinally measured exposure variables \( x_1, x_2, \ldots, x_k \), one time-invariant confounder \( m \), and one distal outcome \( y \)).

2.1. Definitions

We extend the definitions (1-3) provided in Appendix 1 to examine the scenario depicted in Figure 2a.

2.1.1. Definition 4: Standard regression models

Because the relationship between each measurement \( x_i \) and \( y \) is confounded by \( m \) (for \( 1 \leq i \leq k \)), adjustment for \( m \) is necessary to estimate the total effect of \( x_i \) and \( y \) in the standard regression models:

\[
\hat{y}_S^{(1)} = \hat{\alpha}_0^{(1)} + \hat{\alpha}_m^{(1)} m + \hat{\alpha}_{x1}^{(1)} x_1 \\
\hat{y}_S^{(2)} = \hat{\alpha}_0^{(2)} + \hat{\alpha}_m^{(2)} m + \hat{\alpha}_{x1}^{(2)} x_1 + \hat{\alpha}_{x2}^{(2)} x_2 \\
\vdots \\
\hat{y}_S^{(k)} = \hat{\alpha}_0^{(k)} + \hat{\alpha}_m^{(k)} m + \hat{\alpha}_{x1}^{(k)} x_1 + \hat{\alpha}_{x2}^{(k)} x_2 + \cdots + \hat{\alpha}_{xk}^{(k)} x_k.
\]  
(Eq.5)

2.1.2. Definition 5: Unexplained residual (UR) terms

In Figure 2a, it is clear that \( m \) confounds the relationship between \( x_i \) and \( x_1, x_2, \ldots, x_{i-1} \) for \( 2 \leq i \leq k \), and thus adjustment for \( m \) is necessary when regressing \( x_1 \sim x_2, \ldots, x_{i-1} \) to generate each UR term \( e_{xi} \), i.e.:

\[
x_i = \hat{y}_0^{(i)} + \hat{\gamma}_m^{(i)} m + \hat{\gamma}_{x1}^{(i)} x_1 + \hat{\gamma}_{x2}^{(i)} x_2 + \cdots + \hat{\gamma}_{x(i-1)}^{(i)} x_{i-1} + e_{xi}
\]  
(Eq.6)

and

\[
e_{xi} = -\hat{\gamma}_0^{(i)} - \hat{\gamma}_m^{(i)} m - \hat{\gamma}_{x1}^{(i)} x_1 - \hat{\gamma}_{x2}^{(i)} x_2 - \cdots - \hat{\gamma}_{x(i-1)}^{(i)} x_{i-1} + x_i.
\]  
(Eq.7)

In this way, \( e_{xi} \) represents the difference between the actual value of \( x_i \) and the value of \( x_i \) as predicted by all previous measurements \( x_1, x_2, \ldots, x_{i-1} \), **adjusted for the confounding effect of \( m \)**.

2.1.3. Definition 6: Unexplained residuals (UR) models

Furthermore, \( m \) confounds the relationship between \( x_i \) and \( y \), and so adjustment must be made in the composite UR model:

\[
\hat{y}_{UR}^{(k)} = \hat{\beta}_0^{(k)} + \hat{\beta}_m^{(k)} m + \hat{\beta}_{x1}^{(k)} x_1 + \hat{\beta}_{x2}^{(k)} x_2 + \cdots + \hat{\beta}_{exk}^{(k)} e_{xk}.
\]  
(Eq.8)

2.2. Mathematical proofs

The proofs that follow rely upon the following key properties of OLS regression estimators and require the following two lemmas:
**Key properties of OLS estimators:** We may represent the regression equation \( y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon \) in summary notation as:

\[
y = X\beta + \varepsilon,
\]

where: \( y \) represents the vector of \( n \) continuous observations of the outcome; \( X \) represents the \( n \times (k + 1) \) matrix of \( n \) observations for \( k \) continuous covariates and 1 constant; \( \beta \) represents the \( k + 1 \) vector of coefficients for each covariate and constant; and \( \varepsilon \) represents the vector of \( n \) residuals.

The OLS estimate of \( \beta \) is given by:

\[
\hat{\beta} = (X'X)^{-1}X'y.
\]

On the assumption that the inverse matrix exists, this equation has a unique solution.

Further, for the given OLS equation \( y = X\beta + \varepsilon \), it can be shown that the vector of residuals \( (\varepsilon) \) is orthogonal (denoted \( \perp \)) to every column \((1, x_1, x_2, \ldots, x_k)\) of \( X \).

*Note that detailed proofs have not been provided, but can be located in referenced material.*

**Lemma 1:** For two orthogonal components \( \tau \) and \( \delta \) (i.e. \( \tau \perp \delta \)), the estimated coefficients of the regression of \( y \) on \( \tau \) and \( \delta \) are equal to the estimated coefficients for the separate regressions of \( y \) on \( \tau \) and \( y \) on \( \delta \).

**Proof of Lemma 1:** The regression of \( y \) on \( \tau \) and \( \delta \) may be written as:

\[
y = [\tau \quad \delta] [\beta_\tau \quad \beta_\delta] + \varepsilon = \tau \beta_\tau + \delta \beta_\delta + \varepsilon.
\]

From Definition 1, the OLS estimate of \( \beta_\tau \) and \( \beta_\delta \) is given by \( \hat{\beta} = (X'X)^{-1}X'y \). In this scenario,

\[
X'X = \begin{bmatrix} \tau' \ \\
\delta' \end{bmatrix} [\tau \quad \delta] = \begin{bmatrix} \tau' \tau & \tau' \delta \\
\delta' \tau & \delta' \delta \end{bmatrix} = \begin{bmatrix} \tau' \tau & 0 \\
0 & \delta' \delta \end{bmatrix},
\]

where the final equivalency follows from the condition of orthogonality. Then

\[
(X'X)^{-1} = \begin{bmatrix} \tau' \tau & 0 \\
0 & \delta' \delta \end{bmatrix}^{-1} = \begin{bmatrix} (\tau' \tau)^{-1} & 0 \\
0 & (\delta' \delta)^{-1} \end{bmatrix},
\]

and

\[
X'y = \begin{bmatrix} \tau' \ \\
\delta' \end{bmatrix} y = \begin{bmatrix} \tau' y \\
\delta' y \end{bmatrix}.
\]

Combining these elements gives:

\[
\begin{bmatrix} \hat{\beta}_\tau \\
\hat{\beta}_\delta \end{bmatrix} = \begin{bmatrix} (\tau' \tau)^{-1} & 0 \\
0 & (\delta' \delta)^{-1} \end{bmatrix} \begin{bmatrix} \tau' y \\
\delta' y \end{bmatrix} = \begin{bmatrix} (\tau' \tau)^{-1} \tau' y \\
(\delta' \delta)^{-1} \delta' y \end{bmatrix}.
\]

From this, we see that the estimated coefficients are equivalent to those that would be produced for the separate regressions of \( y \) on \( \tau \) and \( y \) on \( \delta \).
Lemma 2: If \( t_i \perp \delta_j \) for \( 0 \leq i \leq h \) and \( 0 \leq j \leq k \), then \( \text{span}(\tau_0, \tau_1, ..., \tau_h) \perp \text{span}(\delta_0, \delta_1, ..., \delta_k) \) for any vectors \( \tau_0, \tau_1, ..., \tau_h, \delta_0, \delta_1, ..., \delta_k \).

Proof of Lemma 2: \( t_i \perp \delta_j \) implies that \( t_i \cdot \delta_j = 0 \) for \( 0 \leq i \leq h \) and \( 0 \leq j \leq k \). Then

\[
\text{span}(\tau_0, \tau_1, ..., \tau_h) \cdot \text{span}(\delta_0, \delta_1, ..., \delta_k) = (c_0 \tau_0 + c_1 \tau_1 + ... + c_h \tau_h) \cdot (d_0 \delta_0 + d_1 \delta_1 + ... + d_k \delta_k)
\]

\[
= c_0 d_0 (\tau_0 \cdot \delta_0) + c_0 d_1 (\tau_0 \cdot \delta_1) + ... + c_0 d_k (\tau_0 \cdot \delta_k) + c_1 d_1 (\tau_1 \cdot \delta_1) + ... + c_1 d_k (\tau_1 \cdot \delta_k) + ...
\]

\[
= 0
\]

Thus, \( \text{span}(\tau_0, \tau_1, ..., \tau_h) \perp \text{span}(\delta_0, \delta_1, ..., \delta_k) \). ■

2.2.1. Covariate orthogonality

We prove that all UR terms \( e_{x_2}, e_{x_3}, ..., e_{x_k} \) are orthogonal to all preceding variables in the composite UR model (Eq.8), and therefore orthogonal to their span; we prove this below.

Lemma 4: \( e_{x_1} \perp e_{x_2}, e_{x_3}, ..., e_{x_{(i-1)}} \), for \( 2 \leq i \leq k \).

Proof of Lemma 4: By construction, \( e_{x_1} \) represents the residuals from the OLS regression of \( x_{i-1} \sim m, x_1, x_2, ..., x_{i-1} \) (Eq.7). Thus, \( e_{x_1} \perp (1, m, x_1, x_2, ..., x_{i-1}) \) from which it follows that \( e_{x_1} \perp \text{span}(1, m, x_1, x_2, ..., x_{i-1}) \) by Lemma 2.

Because \( e_{x_2}, e_{x_3}, ..., e_{x_{(i-1)}} \in \text{span}(1, m, x_1, x_2, ..., x_{i-1}) \) for \( 2 \leq i \leq k \) by construction, we are able to conclude that \( e_{x_1} \perp e_{x_2}, e_{x_3}, ..., e_{x_{(i-1)}} \). ■

Theorem 2: \( e_{x_1} \perp \text{span}(1, m, x_1, e_{x_2}, e_{x_3}, ..., e_{x_{(i-1)}}) \), for \( 2 \leq i \leq k \).

Proof of Theorem 2: \( e_{x_1} \perp 1, m, x_1 \) because \( e_{x_1} \) represents the residuals from the OLS regression of \( x_{i-1} \sim m, x_1, x_2, ..., x_{i-1} \). Further, \( e_{x_1} \perp e_{x_2}, e_{x_3}, ..., e_{x_{(i-1)}} \) for \( 2 \leq i \leq k \) by Lemma 4 above.

Thus, \( e_{x_1} \perp \text{span}(1, m, x_1, e_{x_2}, e_{x_3}, ..., e_{x_{(i-1)}}) \) by Lemma 2. ■

2.2.2. Property (i): \( \hat{Y}_S^{(k)} = \hat{Y}_U^{(k)} \)

Proof of Property (i): As before, this equality follows from the fact that \( \hat{Y}_U^{(k)} \) is a function of the same variables as \( \hat{Y}_S^{(k)} \).

By Definition 6, \( \hat{Y}_U^{(k)} = f(1, m, x_1, e_{x_2}, ..., e_{x_i}) \), where \( e_i = f(1, m, x_1, x_2, ..., x_i) \) by Definition 5.

Thus, it also holds that

\[\text{span}(\tau_0, \tau_1, ..., \tau_h) \perp \text{span}(\delta_0, \delta_1, ..., \delta_k)\]

where the coefficients \( c_0, c_1, c_2, ..., c_k \) are scalars.

---

3 The span of a set of vectors \( \delta_0, \delta_1, \delta_2, ..., \delta_k \) is the set of all possible linear combinations of \( \delta_0, \delta_1, \delta_2, ..., \delta_k \), i.e.

\[\text{span}(\delta_0, \delta_1, \delta_2, ..., \delta_k) = c_0 \delta_0 + c_1 \delta_1 + c_2 \delta_2 + \cdots + c_k \delta_k,\]

where the coefficients \( c_0, c_1, c_2, ..., c_k \) are scalars.
\[ \hat{y}_{UR}^{(i)} = f(1, m, x_1, x_2, \ldots, x_i) . \]

Moreover, by Definition 4,
\[ \hat{y}_{S}^{(i)} = f(1, m, x_1, x_2, \ldots, x_i) . \]

From this, it follows that \( \hat{y}_{S}^{(i)} = \hat{y}_{UR}^{(i)} \) and, consequently, \( \hat{y}_{S}^{(k)} = \hat{y}_{UR}^{(k)} \). ■

### 2.2.3. Property (ii): \( \hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(1)} \)

**Proof of Property (ii):** By definition, \( \hat{y}_{S}^{(i)} = \hat{y}_{UR}^{(i)} = f(1, m, x_1) \), and it is trivially true that \( \hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(1)} \).

Because \( e_{xi} \perp \text{span}(1, m, x_1, e_{x2}, e_{x3}, \ldots, e_{x(i-1)}) \) for \( 2 \leq i \leq k \) by Theorem 2, we conclude that
\[ \hat{\lambda}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(2)} = \ldots = \hat{\lambda}_{x1}^{(k)} \text{ from Lemma 1.} \]

Therefore, \( \hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(1)} \). ■

### 2.2.4. Property (iii): \( \hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{exi}^{(i)} \)

**Proof of Property (iii):** Consider the UR model:
\[ \hat{y}_{UR}^{(i)} = \hat{\alpha}_0^{(i)} + \hat{\lambda}_m^{(i)} m + \hat{\alpha}_{x1}^{(i)} x_1 + \hat{\lambda}_{ex2}^{(i)} x_2 + \ldots + \hat{\lambda}_{exi}^{(i)} e_{xi} . \]

If we substitute the expansion for \( e_{xi} \) (Eq. 7) into this equation and rearrange, we produce:
\[ \hat{y}_{UR}^{(i)} = \hat{\alpha}_0^{(i)} + \hat{\lambda}_m^{(i)} m + \hat{\alpha}_{x1}^{(i)} x_1 + \hat{\lambda}_{ex2}^{(i)} x_2 + \ldots + \hat{\lambda}_{ex(i-1)}^{(i)} x_{i-1} + x_i - f_{x1}^{(i)} m + \hat{\lambda}_{ex2}^{(i)} x_2 + \ldots + \hat{\lambda}_{ex(i-1)}^{(i)} x_{i-1} + x_i - f_{x1}^{(i)} m \]

\[ = \hat{\alpha}_0^{(i)} + \hat{\lambda}_m^{(i)} m + \hat{\lambda}_{ex2}^{(i)} x_2 + \ldots + \hat{\lambda}_{ex(i-1)}^{(i)} x_{i-1} + \hat{\lambda}_{exi}^{(i)} e_{xi} . \]

We have already established that \( \hat{y}_{S}^{(i)} = \hat{y}_{UR}^{(i)} \) (i.e. Property (i)) because they are functions of the same covariates, so it follows that the estimated coefficients for those covariates must themselves be equal. Specifically, we see that the coefficient for \( x_i \) will always equal the coefficient for \( e_{xi} \), i.e.
\[ \hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{exi}^{(i)} . \]

Because \( e_{xi} \perp \text{span}(1, m, x_1, e_{x2}, e_{x3}, \ldots, e_{x(i-1)}) \), we may apply Lemma 1 and conclude that
\[ \hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{ex1}^{(1)} = \hat{\lambda}_{ex2}^{(2)} = \ldots = \hat{\lambda}_{exi}^{(k)} \text{ from which it follows that } \hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{exi}^{(i)} . ■ \]

---

According to the logic applied in this proof may be easily extended to show that \( \hat{\alpha}_m^{(i)} = \hat{\lambda}_m^{(i)} . ■ \)
3. Appendix 3: UR models: Time-varying confounder

Proofs of Properties (i) – (iii) for the scenario depicted in Figure 3a (i.e. \( k \) longitudinally measured exposure variables \( x_1, x_2, \ldots, x_k \), one time-varying confounder \( m_1, m_2, \ldots, m_k \), and one distal outcome \( y \)).

3.1. Definitions

We extend the definitions (1-3) provided in Appendix 1 to examine the scenario depicted in Figure 3a.

3.1.1. Definition 7: Standard regression models

In this scenario, the relationship between each \( x_i \) and \( y \) is confounded by all previous measurements of the exposure \( x_1, x_2, \ldots, x_{i-1} \), as well as all previous and current measurements of the confounder \( m_1, m_2, \ldots, m_i \) (for \( 1 \leq i \leq k \)). These covariates must all be included in the standard regression models to obtain an unbiased estimate of the total causal effect of each measurement \( x_i \) on \( y \), i.e.:

\[
\hat{y}_S^{(1)} = \hat{\alpha}_0^{(1)} + \hat{\alpha}_m^{(1)} m_1 + \hat{\alpha}_x^{(1)} x_1 \\
\hat{y}_S^{(2)} = \hat{\alpha}_0^{(2)} + \hat{\alpha}_m^{(2)} m_1 + \hat{\alpha}_x^{(2)} x_1 + \hat{\alpha}_m^{(2)} m_2 + \hat{\alpha}_x^{(2)} x_2 \\
\vdots \\
\hat{y}_S^{(k)} = \hat{\alpha}_0^{(k)} + \hat{\alpha}_m^{(k)} m_1 + \hat{\alpha}_x^{(k)} x_1 + \cdots + \hat{\alpha}_m^{(k)} m_k + \hat{\alpha}_x^{(k)} x_k. \tag{Eq. 9}
\]

3.1.2. Definition 8: Unexplained residual (UR) terms

The DAG in Figure 3a also makes evident that the relationship between each measurement \( x_i \) and all previous measurements of the exposure \( x_1, x_2, \ldots, x_{i-1} \) is confounded by all previous and current measurements of the confounder \( m_1, m_2, \ldots, m_i \). Thus, we create UR terms \( e_{xi} \) for each measurement of the exposure variable \( x_i \) by adjusting for \( m_1, m_2, \ldots, m_i \), i.e.:

\[
x_i = \hat{y}_0^{(i)} + \hat{\gamma}_m^{(i)} m_1 + \hat{\gamma}_x^{(i)} x_1 + \cdots + \hat{\gamma}_{m(i-1)}^{(i)} m_{i-1} + \hat{\gamma}_{x(i-1)}^{(i)} x_{i-1} + \hat{\gamma}_m^{(i)} m_i + e_{xi} \tag{Eq. 10}
\]

and

\[
e_{xi} = -\hat{\gamma}_0^{(i)} - \hat{\gamma}_m^{(i)} m_1 - \hat{\gamma}_x^{(i)} x_1 - \cdots - \hat{\gamma}_{m(i-1)}^{(i)} m_{i-1} - \hat{\gamma}_{x(i-1)}^{(i)} x_{i-1} - \hat{\gamma}_m^{(i)} m_i + x_i. \tag{Eq. 11}
\]

In this way, \( e_{xi} \) represents the difference between the observed value of \( x_i \) and the value of \( x_i \) as predicted by all previous measurements \( x_1, x_2, \ldots, x_{i-1} \), adjusted for the confounding effects of \( m_1, m_2, \ldots, m_i \).

Previous proofs have relied upon the orthogonality of the terms in the composite UR model (i.e. Theorems 1 and 2 in Appendices 1 and 2, respectively). This necessitates the creation of UR terms \( e_{mi} \) for each measurement of the time-varying confounding variable \( m_i \), for \( 2 \leq i \leq k \). Each \( e_{mi} \) is derived from the OLS regression of \( m_i \) on all previous values of the confounder \( m_1, m_2, \ldots, m_{i-1} \) and all previous values of the exposure \( x_1, x_2, \ldots, x_{i-1} \), i.e.:

\[
m_i = \hat{m}_0^{(i)} + \hat{\eta}_m^{(i)} m_1 + \hat{\eta}_x^{(i)} x_1 + \cdots + \hat{\eta}_{m(i-1)}^{(i)} m_{i-1} + \hat{\eta}_{x(i-1)}^{(i)} x_{i-1} + e_{mi} \tag{Eq. 12}
\]

and
\[ e_{mi} = -\hat{\eta}_0^{(i)} - \hat{\eta}_1^{(i)} m_1 - \hat{\eta}_2^{(i)} x_1 - \cdots - \hat{\eta}_{m(i-1)}^{(i)} m_{i-1} - \hat{\eta}_i^{(i)} x_{i-1} + m_i. \]  
(Eq.13)

These adjustments follow from the DAG in Figure 3a, in which it is evident that \( x_1, x_2, \ldots, x_{i-1} \) confound the relationship between \( m_i \) and \( m_1, m_2, \ldots, m_{i-1} \). Thus, \( e_{mi} \) has a similar interpretation to the original UR terms, in that it represents the part of \( m_i \) unexplained by all previous values \( m_1, m_2, \ldots, m_{i-1} \), adjusted for the confounding effects of \( x_1, x_2, \ldots, x_{i-1} \).

### 3.1.3. Definition 9: Unexplained residuals (UR) models

Finally, we represent the composite UR model as a function of the initial value of the exposure \( x_1 \) and all subsequent URs for the exposure \( e_{x2}, e_{x3}, \ldots, e_{xi} \), and the initial value of the confounder \( m_1 \) and all subsequent URs for the confounder \( e_{m2}, e_{m3}, \ldots, e_{mi} \):

\[ \hat{y}_{UR}^{(k)} = \hat{\lambda}_0^{(k)} + \hat{\lambda}_1^{(k)} m_1 + \hat{\lambda}_x^{(k)} x_1 + \cdots + \hat{\lambda}_e^{(k)} e_{x2} + \cdots + \hat{\lambda}_e^{(k)} e_{xi} + \hat{\lambda}_m^{(k)} m_2 + \cdots + \hat{\lambda}_e^{(k)} e_{mi}. \]

(Eq.14)

### 3.2. Mathematical proofs

The proofs that follow rely upon the following key properties of OLS regression estimators and require the following two lemmas:

**Key properties of OLS estimators:** We may represent the regression equation \( y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon \) in summary notation as:

\[ y = X \beta + \varepsilon, \]

where: \( y \) represents the vector of \( n \) continuous observations of the outcome; \( X \) represents the \( n \times (k + 1) \) matrix of \( n \) observations for \( k \) continuous covariates and 1 constant; \( \beta \) represents the \( k + 1 \) vector of coefficients for each covariate and constant; and \( \varepsilon \) represents the vector of \( n \) residuals.

The OLS estimate of \( \beta \) is given by:

\[ \hat{\beta} = (X'X)^{-1}X'y. \]

On the assumption that the inverse matrix exists, this equation has a unique solution.

Further, for the given OLS equation \( y = X \hat{\beta} + \varepsilon \), it can be shown that the vector of residuals \( (e) \) is orthogonal (denoted \( \perp \)) to every column \((1, x_1, x_2, \ldots, x_k)\) of \( X \).

*Note that detailed proofs have not been provided, but can be located in referenced material.*

**Lemma 1:** For two orthogonal components \( \tau \) and \( \delta \) (i.e. \( \tau \perp \delta \)), the estimated coefficients of the regression of \( y \) on \( \tau \) and \( \delta \) are equal to the estimated coefficients for the separate regressions of \( y \) on \( \tau \) and \( y \) on \( \delta \).

**Proof of Lemma 1:** The regression of \( y \) on \( \tau \) and \( \delta \) may be written as:

\[ y = \begin{bmatrix} \tau & \delta \end{bmatrix} \begin{bmatrix} \beta_\tau \\ \beta_\delta \end{bmatrix} + \varepsilon = \tau \beta_\tau + \delta \beta_\delta + \varepsilon. \]

From Definition 1, the OLS estimate of \( \beta_\tau \) and \( \beta_\delta \) is given by \( \hat{\beta} = (X'X)^{-1}X'y \). In this scenario,

\[ X'X = \begin{bmatrix} \tau' & \delta' \end{bmatrix} \begin{bmatrix} \tau & \delta \end{bmatrix} = \begin{bmatrix} \tau'\tau & \tau'\delta \\ \delta'\tau & \delta'\delta \end{bmatrix} = \begin{bmatrix} \tau'\tau & 0 \\ 0 & \delta'\delta \end{bmatrix}, \]

where the final equivalency follows from the condition of orthogonality. Then
\[
(X'X)^{-1} = \begin{bmatrix}
\tau' \tau & 0 \\
0 & \delta' \delta
\end{bmatrix}^{-1} = \begin{bmatrix}
(\tau' \tau)^{-1} & 0 \\
0 & (\delta' \delta)^{-1}
\end{bmatrix}
\]

and

\[
X'y = \begin{bmatrix}
\tau'^{\dagger} \\
\delta'^{\dagger}
\end{bmatrix} y = \begin{bmatrix}
\tau'y \\
\delta'y
\end{bmatrix}.
\]

Combining these elements gives:

\[
\begin{bmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2
\end{bmatrix} = \begin{bmatrix}
(\tau' \tau)^{-1} & 0 \\
0 & (\delta' \delta)^{-1}
\end{bmatrix} \begin{bmatrix}
\tau'y \\
\delta'y
\end{bmatrix} = \begin{bmatrix}
(\tau' \tau)^{-1} \tau'y \\
(\delta' \delta)^{-1} \delta'y
\end{bmatrix}.
\]

From this, we see that the estimated coefficients are equivalent to those that would be produced for the separate regressions of \(y\) on \(\tau\) and \(y\) on \(\delta\). □

**Lemma 2**: If \(t_i \perp \delta_j\) for \(0 \leq i \leq h\) and \(0 \leq j \leq k\), then \(\text{span}(\tau_0, \tau_1, ..., \tau_h) \perp \text{span}(\delta_0, \delta_1, ..., \delta_k)\) for any vectors \(\tau_0, \tau_1, ..., \tau_h, \delta_0, \delta_1, ..., \delta_k\).

**Proof of Lemma 2**: \(t_i \perp \delta_j\) implies that \(t_i \cdot \delta_j = 0\) for \(0 \leq i \leq h\) and \(0 \leq j \leq k\). Then

\[
\text{span}(\tau_0, \tau_1, ..., \tau_h) \cdot \text{span}(\delta_0, \delta_1, ..., \delta_k) = (c_0 \tau_0 + c_1 \tau_1 + ... + c_h \tau_h) \cdot (d_0 \delta_0 + d_1 \delta_1 + ... + d_k \delta_k)
\]

\[
= c_0 d_0 (\tau_0 \cdot \delta_0) + c_0 d_1 (\tau_0 \cdot \delta_1) + ... + c_0 d_k (\tau_0 \cdot \delta_k) + c_1 d_0 (\tau_1 \cdot \delta_0) + c_1 d_1 (\tau_1 \cdot \delta_1) + ... + c_1 d_k (\tau_1 \cdot \delta_k) + ... + c_h d_0 (\tau_h \cdot \delta_0) + c_h d_1 (\tau_h \cdot \delta_1) + ... + c_h d_k (\tau_h \cdot \delta_k)
\]

\[
= c_0 d_0 (0) + c_0 d_1 (0) + ... + c_0 d_k (0) + c_1 d_0 (0) + c_1 d_1 (0) + ... + c_1 d_k (0) + ... + c_h d_0 (0) + c_h d_1 (0) + ... + c_h d_k (0)
\]

\[
= 0
\]

Thus, \(\text{span}(\tau_0, \tau_1, ..., \tau_h) \perp \text{span}(\delta_0, \delta_1, ..., \delta_k)\). □

### 3.2.1. Covariate orthogonality

Here, we show that: the UR terms for each measurement of the confounder (i.e. \(e_{m_2}, e_{m_3}, ..., e_{m_i}\)) are mutually orthogonal; the UR terms for each measurement of the exposure (i.e. \(e_{x_2}, e_{x_3}, ..., e_{x_i}\)) are mutually orthogonal; and, importantly, the UR terms \(e_{m_2}, e_{m_3}, ..., e_{m_i}\) are orthogonal to \(e_{x_2}, e_{x_3}, ..., e_{x_i}\).

**Lemma 6**: \(e_{m_i} \perp e_{m_2}, e_{m_3}, ..., e_{m(i-1)}\), for \(2 \leq i \leq k\).

**Proof of Lemma 6**: By construction, \(e_{m_i}\) represents the residuals from the OLS regression of \(m_i \sim 1, m_1, x_1, ..., m_{i-1}, x_{i-1}\) (Eq.13). Thus, \(e_{m_i} \perp 1, m_1, x_1, ..., m_{i-1}, x_{i-1}\), which implies \(e_{m_i} \cdot 1 = 0, e_{m_i} \cdot m_1 = 0, e_{m_i} \cdot x_1 = 0, ..., e_{m_i} \cdot m_{i-1} = 0, e_{m_i} \cdot x_{i-1} = 0\).

From this, it follows that \(e_{m_i} \perp \text{span}(1, m_1, x_1, ..., m_{i-1}, x_{i-1})\) from Lemma 2.

---

5 The span of a set of vectors \(\delta_0, \delta_1, ..., \delta_k\) is the set of all possible linear combinations of \(\delta_0, \delta_1, ..., \delta_k\), i.e.:

\[
\text{span}(\delta_0, \delta_1, ..., \delta_k) = c_0 \delta_0 + c_1 \delta_1 + ... + c_k \delta_k,
\]

where the coefficients \(c_0, c_1, ..., c_k\) are scalars.
Because $e_{m2}, e_{m3}, \ldots, e_{m(i-1)} \in \text{span}(1, m_1, x_1, \ldots, m_{i-1}, x_{i-1})$ for $2 \leq i \leq k$ by construction, we are able to conclude that $e_{mi} \perp e_{m2}, e_{m3}, \ldots, e_{m(i-1)}$. 

**Lemma 7:** $e_{xi} \perp e_{x2}, e_{x3}, \ldots, e_{x(i-1)}$, for $2 \leq i \leq k$.

**Proof of Lemma 7:** By construction, $e_{xi}$ represents the residuals from the OLS regression of 

$x_i \sim 1, m_1, x_1, \ldots, m_{i-1}, x_{i-1}, m_i$ (Eq. 12), thus $e_{xi} \perp 1, m_1, x_1, \ldots, m_{i-1}, x_{i-1}, m_i$, which implies $e_{xi} \cdot 1 = 0$, $e_{xi} \cdot m_1 = 0$, $e_{xi} \cdot x_1 = 0$, $\ldots$, $e_{xi} \cdot m_{i-1} = 0$, $e_{xi} \cdot x_{i-1} = 0$, $e_{xi} \cdot m_i = 0$.

From this, it follows that $e_{xi} \perp \text{span}(1, m_1, x_1, \ldots, m_{i-1}, x_{i-1}, m_i)$ from Lemma 2.

Because $e_{x2}, e_{x3}, \ldots, e_{x(i-1)} \in \text{span}(1, m_1, x_1, \ldots, m_{i-1}, x_{i-1}, m_i)$ for $2 \leq i \leq k$ by construction, we are able to conclude that $e_{xi} \perp e_{x2}, e_{x3}, \ldots, e_{x(i-1)}$. 

**Lemma 8:** $e_{xi} \perp e_{mj}$, for $2 \leq i \leq k$ and $2 \leq j \leq k$.

**Proof of Lemma 8:** As established previously, $e_{xi} \perp \text{span}(1, m_1, x_1, \ldots, m_{i-1}, x_{i-1}, m_i)$ by Lemma 2, for $2 \leq i \leq k$. Because $e_{m2}, e_{m3}, \ldots, e_{m(i-1)} \in \text{span}(1, m_1, x_1, \ldots, m_{i-1}, x_{i-1}, m_i)$ by construction, it is evident that $e_{xi} \perp e_{m2}, e_{m3}, \ldots, e_{m(i-1)}$.

Further, $e_{mj} \perp \text{span}(1, m_1, x_1, \ldots, m_{j-1}, x_{j-1})$ by Lemma 2, for $2 \leq j \leq k$. Because $e_{x2}, e_{x3}, \ldots, e_{x(j-1)} \in \text{span}(1, m_1, x_1, \ldots, m_{j-1}, x_{j-1})$ by construction, it is evident that $e_{mj} \perp e_{x2}, e_{x3}, \ldots, e_{x(j-1)}$.

Combining these two results, it follows that $e_{xi} \perp e_{mj}$ for $2 \leq i \leq k$ and $2 \leq j \leq k$.

**Theorem 3:** $\text{span}(e_{xi}, e_{mi}) \perp \text{span}(1, m_1, x_1, \ldots, e_{m(i-1)}, e_{x(i-1)})$, for $2 \leq i \leq k$.

**Proof of Theorem 3:** By definition, $e_{xi} \perp 1, m_1, x_1$. As established in Lemmas 7 and 8, $e_{xi} \perp e_{x2}, e_{x3}, \ldots, e_{x(i-1)}, e_{m1}, \ldots, e_{m(i-1)}$.

Further, $e_{mi} \perp 1, m_1, x_1$ by definition, and as established in Lemmas 6 and 8, $e_{mi} \perp e_{x2}, e_{x3}, \ldots, e_{x(i-1)}, e_{m1}, \ldots, e_{m(i-1)}$.

Thus, by Lemma 2, it follows that $\text{span}(e_{xi}, e_{mi}) \perp \text{span}(1, m_1, x_1, \ldots, e_{m(i-1)}, e_{x(i-1)})$.

3.2.2. Property (i): $\hat{y}_S^{(k)} = \hat{y}_S^{(k)}$

**Proof of Property (i):** As previously, Property (i) follows from the fact that $\hat{y}_S^{(k)}$ is a function of the same variables as $\hat{y}_S^{(i)}$.

By Definition 9, $\hat{y}_S^{(i)} = f(1, m_1, x_1, e_{m2}, e_{x2}, \ldots, e_{mi}, e_{xi})$, where $e_{xi} = f(1, m_1, x_1, \ldots, m_i, x_i)$ and $e_{mi} = f(1, m_1, x_1, \ldots, m_{i-1}, x_{i-1}, m_i)$ by Definition 8. Thus, it also holds that 

$\hat{y}_S^{(i)} = f(1, m_1, x_1, \ldots, m_i, x_i)$.

Moreover, by Definition 7, 

$\hat{y}_S^{(k)} = f(1, m_1, x_1, \ldots, m_i, x_i)$.

From this, it follows that $\hat{y}_S^{(i)} = \hat{y}_S^{(i)}$ and, consequently, $\hat{y}_S^{(k)} = \hat{y}_S^{(k)}$.
3.2.3. Property (ii): $\hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(k)}$

Proof of Property (ii): By definition, $\hat{y}_S^{(1)} = y_{UR}^{(1)} = f(1, m_1, x_1)$, and it is trivially true that $\hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(k)}$.

Because $\text{span}(e_{x1}, e_{m1}) \perp \text{span}(1, m_1, x_1, \ldots, e_{m(i-1)}, e_{x(i-1)})$ for $2 \leq i \leq k$ by Theorem 3, we are able to conclude that $\hat{\lambda}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(2)} = \cdots = \hat{\lambda}_{x1}^{(k)}$ by applying Lemma 1.

Therefore, $\hat{\alpha}_{x1}^{(1)} = \hat{\lambda}_{x1}^{(k)}$.  

3.2.4. Property (iii): $\hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{exi}^{(k)}$

Proof of Property (iii): Consider the UR model:

$$y_{UR}^{(i)} = \hat{\alpha}_{0}^{(i)} + \hat{\alpha}_{1}^{(i)} x_1 + \hat{\alpha}_{m1}^{(i)} m_1 + \hat{\alpha}_{x2}^{(i)} x_2 + \hat{\alpha}_{m2}^{(i)} m_2 + \hat{\alpha}_{ex}^{(i)} e_{x1} + \hat{\alpha}_{em}^{(i)} e_{m1} + \hat{\alpha}_{ex}^{(i)} e_{x1} + \hat{\alpha}_{em}^{(i)} e_{m1} + \hat{\alpha}_{x}^{(i)} e_{x1} + \hat{\alpha}_{m}^{(i)} e_{m1}.$$

By substituting the expansions for $e_{x1}$ (Eq.11) and $e_{m1}$ (Eq.13) into this equation and rearranging, we produce:

$$y_{UR}^{(i)} = \hat{\alpha}_{0}^{(i)} + \hat{\alpha}_{1}^{(i)} x_1 + \hat{\alpha}_{m1}^{(i)} m_1 + \hat{\alpha}_{x2}^{(i)} x_2 + \hat{\alpha}_{m2}^{(i)} m_2 + \hat{\alpha}_{ex}^{(i)} e_{x1} + \hat{\alpha}_{em}^{(i)} e_{m1} + \hat{\alpha}_{x}^{(i)} e_{x1} + \hat{\alpha}_{m}^{(i)} e_{m1}.$$

Having established that $y_{S}^{(i)} = y_{UR}^{(i)}$ (i.e. Property (i)) because they are functions of the same covariates, it follows that the estimated coefficients for those covariates must themselves be equal.

Specifically, we see that the coefficient for $x_1$ will always equal the coefficient for $e_{x1}$, i.e. $\hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{exi}^{(k)}$.

Finally, using the fact that $e_{x1} \perp \text{span}(1, m_1, x_1, e_{m2}, e_{x2}, \ldots, e_{m(i-1)}, e_{x(i-1)}, e_{m1})$, we apply Lemma 1 and conclude that $\hat{\lambda}_{exi}^{(1)} = \hat{\lambda}_{exi}^{(2)} = \cdots = \hat{\lambda}_{exi}^{(k)}$, from which it follows that $\hat{\alpha}_{x1}^{(i)} = \hat{\lambda}_{exi}^{(k)}$.

Although no causal meaning/significance can be attributed to the intercept term or the coefficients of the UR terms for the confounder $e_{m2}, \ldots, e_{m1}$, the logic applied in this proof may be easily extended to show that $\hat{\alpha}_0^{(i)} = \hat{\lambda}_0^{(k)}$ and $\hat{\alpha}_{m1}^{(i)} = \hat{\lambda}_{em1}^{(i)}$, $\hat{\alpha}_{m2}^{(i)} = \hat{\lambda}_{em2}^{(i)}$, $\hat{\alpha}_{m3}^{(i)} = \hat{\lambda}_{em3}^{(i)}$, respectively.
4. Appendix 4: Details of standard error simulation

4.1. DAG

Path coefficients represent bivariate correlations.

4.2. Correlation matrix based upon DAG

<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.40</td>
<td>1.00</td>
<td>-</td>
</tr>
<tr>
<td>$y$</td>
<td>-0.22</td>
<td>-0.34</td>
<td>1.00</td>
</tr>
</tbody>
</table>

4.3. Population parameters used in simulation

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>10.00</td>
<td>2.50</td>
</tr>
<tr>
<td>$x_2$</td>
<td>15.00</td>
<td>3.75</td>
</tr>
<tr>
<td>$y$</td>
<td>20.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>

4.4. Annotated R code

```r
# load packages required for simulation
require(Matrix); require(matrixcalc); require(MASS); require(dagitty); require(devtools)
# devtools::install_github("jtextor/dagitty/r") # update regularly

# load packages required for simulation
require(Matrix); require(matrixcalc); require(MASS); require(dagitty); require(devtools)
# devtools::install_github("jtextor/dagitty/r") # update regularly
```
# converts SDs and pairwise correlations to a covariance matrix

Covar <- function(n=2, SD=data.frame(1,1), c.vec=data.frame(0.5)) {
  check   <- n-length(SD)
  if (check !=0) stop("Incorrect SD specifications!")
  check   <- (n*(n-1)/2)-length(c.vec)
  if (check !=0) stop("Incorrect correlation specifications!")
  Cor     <- NULL
  for (i in 1:(n+1)) {
    Row <- NULL
    for (j in 1:(n+1)) {
      if (i==j) Element <- 1
      else if (i<j) Element <- c.vec[((i-1)*(2*n-i)/2)+(j-i)]
      else if (i>j) Element <- c.vec[((j-1)*(2*n-j)/2)+(i-j)]
    Row <- c(Row,Element)
  }
  Cor <- rbind(Cor,Row)
} # cov(i,j) = cor(i,j)*sd(i)*sd(j)

Cov <- matrix(nrow=n,ncol=n)
for (i in 1:n) { for (j in 1:n) { Cov[i,j] <- Cor[i,j]*SD[i]*SD[j] }}

Cov <- as.matrix(forceSymmetric(Cov))
if (!is.positive.definite(Cov)) {
  print("Warning: covariance matrix made Positive Definite")
  Cov <- as.matrix(nearPD(Cov)$mat) }

return(Cov)
}

#########
## DAG ##
#########

dag1 <- dagitty('dag{
  X1 [pos="0.2,0.2"]
  X2 [pos="0.6,0.2"]
  Y [pos="1,1"]
  X1 -> X2 [beta=0.4]
  X1 -> Y [beta=-0.1]
  X2 -> Y [beta=-0.3]
}
)'

plot(dag1)
mod <- lm(Y~X1+X2, data=simulateSEM(dag1, empirical=TRUE))

#############################################################
## COVARIANCE MATRIX ##
#############################################################

MyData  <- simulateSEM(dag1, empirical=TRUE)  # standardised data
Names   <- c("X1","X2","Y")
SetCor  <- cor(MyData); Corr <- SetCor[lower.tri(SetCor)]
N       <- 1000
X1.mu <- 10
X2.mu <- 15
Y.mu <- 20
Mu <- c(X1.mu, X2.mu, Y.mu)
X1.sd <- X1.mu/4
X2.sd <- X2.mu/4
Y.sd <- Y.mu/4
SD <- c(X1.sd, X2.sd, Y.sd)
MyCov <- Covar(3, SD, Corr)

############################
## SIMULATION ##
############################

# set storage for SEs for X1
seX1.reg <- NULL  # standard regression models
seX1.UR <- NULL  # UR models (as reported)
seX1.UR.boot <- NULL  # UR models (bootstrapped)

# set storage for SEs for X2/e2
seX2.reg <- NULL  # standard regression models
see2.UR <- NULL  # UR models (as reported)
see2.UR.boot <- NULL  # UR models (bootstrapped)

set.seed(23)
for (i in 1:1000) {
  # simulate N observations
  MyData  <- data.frame(mvrnorm(N, Mu, MyCov, empirical=FALSE)); names(MyData) <- Names

  # create standard regression model for X1 and save SE
  modX1 <- lm(Y~X1, data=MyData); seX1.reg <- c(seX1.reg, summary(modX1)$coefficients[2,2])

  # create standard regression model for X2 and save SE
  modX2 <- lm(Y~X1+X2, data=MyData); seX2.reg <- c(seX2.reg, summary(modX2)$coefficients[3,2])

  # create UR term
  modX2.resid <- lm(X2~X1, data=MyData); MyData$e2 <- modX2.resid$residuals

  # create UR model and save SEs for coeffs
  modUR <- lm(Y~X1+e2, data=MyData)
  seX1.UR <- c(seX1.UR, summary(modUR)$coefficients[2,2])
  see2.UR <- c(see2.UR, summary(modUR)$coefficients[3,2])

  # use bootstrapping to create distribution of coefficients for UR model
  coeffX1.UR.boot <- NULL  # set storage for coeffs for X1 from UR model
  coeffe2.UR.boot <- NULL  # set storage for coeffs for e2 from UR model

  for (j in 1:1000) {
    select <- sample(c(1:1000), 1000, replace=TRUE)
```r
# create UR term
modX2.resid.boot <- lm(X2~X1, data=MyData.boot); MyData.boot$e2 <-
modX2.resid.boot$residuals

# create UR models and save coeffs
modUR.boot <- lm(Y~X1+e2, data=MyData.boot)
coeffX1.UR.boot <- c(coeffX1.UR.boot, summary(modUR.boot)$coefficients[2,1])
coeffe2.UR.boot <- c(coeffe2.UR.boot, summary(modUR.boot)$coefficients[3,1])
}

# calculate SES for UR model as standard deviation of distribution of coefficients
seX1.UR.boot <- c(seX1.UR.boot, sd(coeffX1.UR.boot))
see2.UR.boot <- c(see2.UR.boot, sd(coeffe2.UR.boot))

# load required packages, import fonts
require(ggplot2); require(gridExtra); require(extrafont); require(Hmisc)
font_import(pattern="[C/c]alibri"); loadfonts(device="win") ## use fonttable() to see options

# function to produce summary statistics (mean and +/- sd)
data_summary <- function(x) {
  m <- mean(x)
ymin <- m - sd(x)
ymax <- m + sd(x)
  return(c(y=m, ymin=ymin, ymax=ymax))
}

# create stacked data frames for each pairwise comparison
DataFrameX1 <- stack(data.frame(seX1.reg,seX1.UR,seX1.UR.boot))
DataFrameX2 <- stack(data.frame(seX2.reg,see2.UR,see2.UR.boot))

# X1 plot
plotX1 <- ggplot(DataFrameX1, aes(x=ind, y=values)) +
  geom_violin(fill="gray60", color="gray30", size=1.2, trim=TRUE) +
  stat_summary(fun.data=data_summary, color="gray90", size=0.7) +
  scale_x_discrete(name="", labels=c("Standard \nregression \nmodels","Unexplained \nresiduals 
models \n(reported)", "Unexplained \nresiduals 
models \n(bootstrapped)")) +
  scale_y_continuous(name="Standard error") +
  ggtitle("Exposure: x1") +
  theme_bw() +
  theme(axis.line=element_line(size=1, colour="black"),
  panel.border=element_blank(),
  #panel.grid.major=element_blank(),
  panel.grid.minor=element_blank(),
  plot.title=element_text(size=16, hjust = 0.5, family="Calibri"),
```

```r
# X1 plot
plotX1 <- ggplot(DataFrameX1, aes(x=ind, y=values)) +
  geom_violin(fill="gray60", color="gray30", size=1.2, trim=TRUE) +
  stat_summary(fun.data=data_summary, color="gray90", size=0.7) +
  scale_x_discrete(limits=c("seX1.reg", "see1.UR", "see1.UR.boot"), name="", labels=c("Standard 
regression \nmodels","Unexplained \nresiduals \nmodels \nreported","Unexplained \nresiduals 
models \n(bootstrapped)")) +
  scale_y_continuous(name="Standard error") +
  ggtitle("Exposure: x1") +
  theme_bw() +
  theme(axis.line=element_line(size=1, colour="black"),
        panel.border=element_blank(),
        panel.grid.major=element_blank(),
        panel.grid.minor=element_blank(),
        plot.title=element_text(size=16, hjust = 0.5, family="Calibri"),
        text=element_text(size=13, family="Calibri Light"),
        axis.text.x=element_text(size=13),
        axis.text.y=element_text(size=11),
        plot.margin=unit(c(0.5,0.5,0.5,0.5),"cm"),
        legend.position="none")

# X2 plot
plotX2 <- ggplot(DataFrameX2, aes(x=ind, y=values)) +
  geom_violin(fill="gray60", color="gray30", size=1.2, trim=TRUE) +
  stat_summary(fun.data=data_summary, color="gray90", size=0.7) +
  scale_x_discrete(limits=c("seX2.reg", "see2.UR", "see2.UR.boot"), name="", labels=c("Standard 
regression \nmodels","Unexplained \nresiduals \nmodels \nreported","Unexplained \nresiduals 
models \n(bootstrapped)")) +
  scale_y_continuous(name="Standard error") +
  ggtitle("Exposure: x2") +
  theme_bw() +
  theme(axis.line=element_line(size=1, colour="black"),
        panel.border=element_blank(),
        panel.grid.major=element_blank(),
        panel.grid.minor=element_blank(),
        plot.title=element_text(size=16, hjust = 0.5, family="Calibri"),
        text=element_text(size=13, family="Calibri Light"),
        axis.text.x=element_text(size=13),
        axis.text.y=element_text(size=11),
        plot.margin=unit(c(0.5,0.5,0.5,0.5),"cm"),
        legend.position="none")

# composite plot
composite <- grid.arrange(plotX1, plotX2,
                            ncol=2, nrow=1,
                            widths=c(5,5), heights=8)
```
References


Figure 1:
(a) Nonparametric causal diagram (DAG) representing the hypothesised data-generating process for k longitudinal measurements of exposure x (i.e. \(x_1, x_2, ..., x_k\)) and one distal outcome y. The terms \(e_{x1}, ..., e_{xk}\) and \(e_y\) represent all unexplained causes of \(x_1, ..., x_k\) and y, respectively, and are included to explicitly reflect uncertainty in all endogenous nodes (whether modelled or not).
(b) Path diagrams depicting the k standard regression models that would be constructed to estimate the total causal effect of each of \(x_1, x_2, ..., x_k\) on y (i.e. Eq.5). For each model, only the final coefficient may be interpreted as a total causal effect; all other coefficients are greyed to illustrate that no such interpretation should be made for them.
(c) Path diagrams depicting the UR model, consisting of \(k - 1\) preparation regressions (i.e. Eq.6) and a final composite regression model (i.e. Eq.7, with \(i = k\)).
Figure 2:
(a) Nonparametric causal diagram (DAG) representing the hypothesised data-generating process for $k$ longitudinal measurements of exposure $x$ (i.e. $x_1, x_2, ..., x_k$), one distal outcome $y$, and one time-invariant confounder $m$. The terms $\epsilon_m, \epsilon_{x1}, ..., \epsilon_{xk}$ and $\epsilon_y$ represent all unexplained causes of $m, x_1, ..., x_k$, and $y$, respectively, and are included to explicitly reflect uncertainty in all endogenous nodes (whether modelled or not).

(b) Path diagrams depicting the $k$ standard regression models that would be constructed to estimate the total causal effect of each of $x_1, x_2, ..., x_k$ on $y$ (i.e. Eq.9). For each model, only the final coefficient may be interpreted as a total causal effect; all other coefficients are greyed to illustrate that no such interpretation should be made for them.

(c) Path diagrams depicting the UR model, consisting of $k - 1$ preparation regressions (i.e. Eq.10) and a final composite regression model (i.e. Eq.11, with $i = k$).
Figure 3:
(a) Nonparametric causal diagram (DAG) representing the hypothesised data-generating process for $k$ longitudinal measurements of exposure $x$ (i.e. $x_1, x_2, ..., x_k$), one distal outcome $y$, and $k$ longitudinal measurements of one time-varying confounder $m_1, m_2, ..., m_k$. The terms $e_{m1}, e_{m2}, ..., e_{mk}$ and $e_y$ represent all unexplained causes of $m_1, m_2, x_1, x_2, ..., x_k$, and $y$, respectively, and are included to explicitly reflect uncertainty in all endogenous nodes (whether modelled or not).
(b) Path diagrams depicting the $k$ standard regression models that would be constructed to estimate the total causal effect of each of $x_1, x_2, ..., x_k$ on $y$ (i.e. Eq.12). For each model, only the final coefficient may be interpreted as a total causal effect; all other coefficients are greyed to illustrate that no such interpretation should be made for them.
(c) Path diagrams depicting the UR model, consisting of $2(k - 1)$ preparation regressions (i.e. Eq.13 and Eq.14) and a final composite regression model (i.e. Eq.15, with $i = k$).
Figure 4: Violin plots comparing the standard errors associated with equivalent coefficients estimated in standard regression vs. UR models, for data simulated based upon the scenario depicted in Figure 1a (with $k = 2$). Horizontal bars within each distribution represent the mean ± 1 standard deviation.