WAGE EFFECTS OF DRINKING AND SMOKING: 
AN ANALYSIS USING AUSTRALIAN TWINS DATA

by

Yew Liang Lee

DISCUSSION PAPER 99.22

DEPARTMENT OF ECONOMICS
THE UNIVERSITY OF WESTERN AUSTRALIA
NEDLANDS, WESTERN AUSTRALIA 6907
WAGE EFFECTS OF DRINKING AND SMOKING:
AN ANALYSIS USING AUSTRALIAN TWINS DATA

by

Yew Liang Lee

Department of Economics
The University of Western Australia

DISCUSSION PAPER 99.22

DEPARTMENT OF ECONOMICS
THE UNIVERSITY OF WESTERN AUSTRALIA
NEDLANDS, WESTERN AUSTRALIA  6907

ISSN 0811-6067
ISBN 0-86422-948-8
WAGE EFFECTS OF DRINKING AND SMOKING: AN ANALYSIS USING AUSTRALIAN TWINS DATA*

by

YEW LIANG LEE

Department of Economics,

University of Western Australia

This paper extends the studies by Ashenfelter and Krueger (1994) and Levine et al. (1997) to analyse the impacts of genetic endowments and shared family background in the determination of the wage effects of drinking and smoking for the Australian labour market. Estimates from alternative specifications indicate that drinking leads to an increase in earnings relative to abstention. Both past and present smoking habits have negative impacts on earnings. Family background is positively related to current smoking behaviour and negatively related to drinking behaviour. Greater genetic endowments reduce the likelihood of smoking but increase the drinking behaviour.

* The author would like to express his sincere gratitude to Nick Martin (Queensland Institute of Medical Research) for the provision of the Australian Twin Registry Data and to Paul Miller and W. Lee Hansen for helpful comments. Financial support from an Overseas Postgraduate Research Scholarship is also gratefully acknowledged.
I Introduction

Given a well-functioning labour market, more productive workers will earn more. Most studies of wage determination have focused on factors that enhance productivity, such as education and labour market experience. Recently, a number of studies have examined factors that might reduce productivity, such as drinking (e.g., Berger and Leigh (1988), Mullahy and Sindelar (1993), Kenkel and Ribar (1994), Heien (1996), Hamilton and Hamilton (1997) and Zarkin et al. (1998)) and smoking (e.g., Levine et al. (1997)). Alcohol consumption and cigarette smoking may have both short-run and long-run effects on productivity. It is often argued that short-run drinking and smoking problems are linked to workplace measures such as job withdrawal behaviours (e.g., taking long lunch breaks), high absenteeism and low productivity. In the long run, both drinking and smoking problems may reduce productivity and wages indirectly through the worker’s health capital, schooling capital, and labour market experience.

The studies of wage effects of drinking and smoking are important because they allow cost-benefit analyses of addictive behaviour to be conducted. Alchim (1995) points out that the benefits associated with smoking accrue only to the individual smoker, and therefore, are private benefits. These benefits include pleasure of smoking and a relaxing sensation. Smoking also enhances acceptance by their peers for some people. In the case of addiction, smoking prevents the discomfort that may arise from the attempt to quit smoking. Turner et al. (1981) point out that moderate alcohol consumption leads to a lower risk of coronary heart disease relative to abstainers.
In terms of costs, Alchin (1995) argues that smoking imposes health costs not only on individuals but also on society as a whole. The total cost of smoking includes its private and social costs. The private costs of smoking are incurred by the smokers themselves. They include the cost of buying cigarettes, the cost of private medical treatment for smoking-related sickness, and the lower wages associated with any reductions in productivity among drinkers and smokers. On the other hand, other social costs of smoking are incurred by non-smokers who bear the cost of adverse health effects due to passive smoking and may subsidise medical treatment through public health systems. The types of private and social costs identified for smoking generally are also applicable in the case of drinking. One of the main social costs of alcohol abuse is the costs associated with car accidents.

Thus, it is apparent that any evaluation of the net costs of addictive behaviour will require information on the wage effects of drinking and smoking. These wage effects must be determined after making allowance for other influences on wages. In this regard, the recent work by Ashenfelter and Krueger (1994) provides the starting point for the current study. They renewed interest in the use of data on twins as a means of obtaining estimates of the returns to schooling that were net of the influence of genetics and shared family background, factors that are held to bias estimates of the returns to schooling in studies based on samples of individuals. In this analysis, the work of Ashenfelter and Krueger (1994) is extended to consider the impact of genetics and shared family background in the determination of the wage effects of addictive behaviours, namely drinking and smoking. This analysis is, therefore, also an extension of Levine et al.’s (1997) work where the effects of smoking on wages were studied using both data on siblings and on individuals over time. It also provides the first evidence for these wage effects for the Australian labour market.
This paper is organised as follows. Section II outlines the theoretical framework of the analysis proposed by Ashenfelter and Krueger (1994). Three models (namely, the conventional model, the fixed-effects model and the selection-effects model) are employed in this study. Section III presents a description of the Australian Twin Registry data. Empirical results of the models using the Australian twins data are reported in section IV. Section V provides a summary and conclusion.

II Theoretical Framework

The model of Ashenfelter and Krueger (1994) is a generalisation of Taubman (1976). Their model of earnings determination may be written as

\[ y_j = \alpha X_j + \beta Z_j + \mu_j + G_j + \varepsilon_j, \]

where \( X_j \) represents the observed components of family background, \( Z_j \) is a vector of individual characteristics that includes variables like education level\(^1\), marital status, and drinking and smoking behaviours, \( \mu_j \) is an unobserved component of family background, \( G_j \) is genetic endowments and \( \varepsilon_j \) is a random error term. The \( X_j \) and \( \mu_j \) terms offer a more detailed representation of the broad family effects (\( N_j \)) considered by Taubman (1976).

In most studies based on equation (1), researchers have difficulty obtaining useable data on measures of \( X \), \( \mu \) and \( G \). The estimates of the returns to schooling (and of the

---

\(^1\) In equation (1), \( Z \) is defined to include a constant term.
wage effects of drinking and smoking) have therefore been characterised by omitted-variables bias. Taubman (1976) employed data on twins to circumvent this problem.

Denoting the logarithmic earnings of twins 1 and 2 in family \( i \) as \( y_{1i} \) and \( y_{2i} \), respectively, their earnings functions are given as

\[
(2) \quad y_{1i} = \alpha X_{1i} + \beta Z_{1i} + \mu_i + G_{1i} + \varepsilon_{1i},
\]

and

\[
(3) \quad y_{2i} = \alpha X_{2i} + \beta Z_{2i} + \mu_i + G_{2i} + \varepsilon_{2i},
\]

where the two equations are assumed to be identical for the two twins.

In the "within-twins" estimation used in this line of research, the difference in the earnings of the members of a twin pair can be represented by the difference between equations (2) and (3), and this may be expressed as:

\[
(4) \quad \Delta y_i = \alpha \Delta X_i + \beta \Delta Z_i + \Delta \mu_i + \Delta G_i + \Delta \varepsilon_i,
\]

where \( \Delta \) denotes the within-pair difference. This is known as the "fixed-effects" estimation.

MZ twins have, by definition, identical genetic composition.\(^2\) If reared together, they also share the same family background. Therefore, for MZ twins reared together,

\(^2\) There are two types of twins, namely monozygotic (MZ) and dizygotic (DZ). MZ twins, also known as identical twins, are genetically identical and have the same innate ability. They are the result of the splitting of an already fertilised egg. DZ or fraternal twins do not share the same genetic composition as they are the result of two different eggs fertilised by two different sperm. They share, on average, one-half of their genes and therefore are no more alike than randomly drawn siblings.
\[ \Delta X_i = \Delta \mu_i = \Delta G_i = 0 \] in the fixed-effects version of the equation. Hence, for MZ twins, equation (4) becomes

\[
(5) \quad \Delta y_i = \beta \Delta Z_i + \Delta \varepsilon_i.
\]

For DZ twins reared together, however, while \( \Delta X_i = \Delta \mu_i = 0, \Delta G_i \neq 0 \), as DZ twins are no more alike than ordinary siblings. Hence, for DZ twins, equation (4) becomes

\[
(6) \quad \Delta y_i = \beta \Delta Z_i + \Delta G_i + \Delta \varepsilon_i.
\]

Estimation of equation (6) on a sample of DZ twins reared together will result in estimates of \( \beta \) that are free of the bias associated with omitted family effects. The estimates will, however, be subject to bias associated with the omitted genetic factors. In studies such as Taubman (1976) and Ashenfelter and Krueger (1994), estimates obtained from samples of individuals (using equation (1)), of MZ twins (using equation (5)) and of DZ twins (using equation (6)) are compared to assess the extent of bias in the estimates of \( \beta \) due to omitted family and genetic effects (from the comparison of MZ twins and individuals) and due to omitted genetic effects (from the comparison of MZ and DZ twins).

An alternative to the fixed-effects model is the selection-effects model. The selection-effects model was introduced by Ashenfelter and Krueger (1994) and provides an explicit account of the family effects in the earnings equation. In this model, the unobservable family effects are treated as being dependent on a number of
observables. This gives the following expression for the correlation between the family effect and observables

\[(7) \quad \mu_i = \gamma Z_{1i} + \gamma Z_{2i} + \delta X_i + \omega_i,\]

where \(\omega_i\) is assumed to be uncorrelated with \(Z_{1i}, Z_{2i},\) and \(X_i\). A further assumption is that correlations between the family effect and the observables for each twin are the same. The coefficients \(\gamma\) measure the “selection effect” which relates earnings and the observables. The availability of twins data makes the measurement of this selection effect possible.

To obtain the reduced form of this model, equation (7) is substituted into both equations (2) and (3). This gives

\[(8) \quad y_{1i} = [\alpha + \gamma] X_i + [\beta + \gamma] Z_{1i} + G_{1i} + \varepsilon_{1i},\]

and

\[(9) \quad y_{2i} = [\alpha + \gamma] X_i + [\beta + \gamma] Z_{2i} + G_{2i} + \varepsilon_{2i},\]

where \(\varepsilon_{1i} = \omega_i + \varepsilon_{1i}\) and \(\varepsilon_{2i} = \omega_i + \varepsilon_{2i}\). In equations (8) and (9), it can be seen that \(Z_{1i}\) and \(Z_{2i}\) enter into each sibling’s earnings equation. This is a result of the correlation between the family effect and any observable that varies across twins, which arises from selection effects. The coefficients, \(\beta\), measure the structural effect of any observable that varies across twins on earnings. The possibility that the coefficients \(\gamma\) may be estimated allows the coefficients of the variables that vary across twins, \(\beta\), to be estimated too.
The contribution of the current study is the development of the range of influences captured by the Z term in the earnings equation. In the work by Ashenfelter and Krueger (1994) for the U.S. labour market, and Miller et al. (1995) for the Australian labour market, the main wage determinant in Z was the level of schooling. Here we also consider drinking and smoking behaviour.

The relationships between alcohol use and earnings and between smoking and earnings are receiving greater attention in the literature. Completing one of the first studies on the relationship between alcohol use and wages, Berger and Leigh (1988) find that drinkers earn significantly higher wages than non-drinkers. Hamilton and Hamilton (1997) report that moderate drinkers (72 percent of the sample) earn 7 percent more than non-drinkers (18 percent of the sample) while heavy drinkers (10 percent of the sample) earn 14 percent more than non-drinkers. However, in a study that takes a wider view of the labour market consequence of drinking, Mullahy and Sindelar (1993) find that alcohol use may affect income more by reducing the likelihood of labour market participation than by affecting the wages of workers.

While acknowledging that alcohol use may affect a wide range of labour market phenomena, this study will be restricted, by nature of the data available, in its focus to the relationship between drinking and the measures of income reviewed in Appendix A.

Levine et al. (1997) employ the National Longitudinal Survey of Youth (NLSY) data to analyse the effects of cigarette smoking on wages. They use a variety of approaches, including study of individuals at a point in time, study of individuals over time, and study of siblings. The analysis using siblings allows the effects of smoking behaviour on wages or employment, net of differences in unobservable family
characteristics, to be measured. Levine et al. (1997) report that smokers earn about 4 to 7 percent less than non-smokers.

III The Data and Sample

This analysis uses data from the Australian Twin Registry which were gathered in two surveys, in 1980-1982 and 1988-1989. A mail-out survey was conducted in 1980-1982 of all 5967 adult twin pairs throughout Australia aged over 18 years who were enrolled in the Australian National Health and Medical Research Council Twin Registry during that period. Being enrolled in this registry and taking part in the survey were both voluntary. The response rate was about 64 percent, with replies being received from 3808 complete pairs of twins after one or two reminders to non-respondents.

The sample of 3808 complete twin pairs was followed up in 1988-1989, and responses were obtained from 6327 individuals, which includes 2995 complete pairs of twins.

The data utilised in this study are mostly from the follow-up survey. Information on how often the members of twin pairs contacted and saw each other during childhood and during the eight-year period prior to the survey was recorded. Other critical information collected includes the respondent’s family background, socioeconomic status and personal details, personality, and feelings and attitude. Data on the respondent’s family background include extensive information on the respondent’s parents, siblings, children, and marital status. Education level, employment status, income level, and occupation form the major measures of socioeconomic status.
Information on personal details was derived from questions asked about body size, general health, and drinking and smoking habits. Each respondent was asked not only to report the highest level of education he/she completed, but also to report that of his/her twin, parents and spouse. As a result of this information, estimates of the reliability of the education measures are possible.

As pointed out in Miller et al. (1995), the twins in the survey have, on average, about one year of schooling more than the national average recorded in the 1986 Australian Census of Population and Housing. Furthermore, the twins are, on average, one year younger than the population of 20-64 year olds. It is generally found that samples of twins have these features compared to national data (see, for example, Ashenfelter and Krueger (1994)).

It is crucial to note that some of the information used in this study was gathered in categorical form. The data on education were collected in seven categories. Also collected in categorical form are the data on income, with only eight categories being used. Due to the broad intervals specified in most of the categories (especially those in the middle of the income distribution), more than one-third of the twins who were employed had twins who were in the same income category. This gives rise to an inadequacy of the income data for many economic applications. In place of the categorical income data, a measure of the mean earnings of the occupation of employment is used. Other studies that have used this approach include Griliches (1977), Nickell (1982), and Behrman et al. (1994). Details on the method of construction are set out in Appendix A.

Of the 2995 complete pairs who responded to the 1988-1989 survey, there were 2943 pairs with educational attainment available from both the 1980-1982 and 1988-

---

3 The seven categories are: < 7 years of schooling; 11-12 years of schooling; apprenticeship, diploma, certificate; technical or teachers college; university, first degree; university, postgraduate degree.
1989 surveys. In order to be included in the sample used for this study, the twins had to be between 20 and 64 years old. Since 170 pairs of twins were aged above 64 years, they were eliminated from the 2943 pairs of twins who responded to the 1988-1989 survey. 351 pairs were excluded due to the lack of information on the co-twin’s educational attainment. There was a further loss of 580 pairs of twins as a consequence of missing data on occupational status. One of the requirements of this study is that both members of any twin pair have to engage in either full-time or part-time employment. However, 672 twin pairs did not meet this criterion, thus resulting in 1170 pairs.

Taking account of the drinking habits of the twins reduces the number of twin pairs to 1160 pairs, since 10 pairs of twins failed to report their drinking habits. A further reduction of 5 twin pairs results due to the lack of information on the smoking habits of the twins. As a consequence, this analysis is based on a total of 1155 complete pairs of twins (596 pairs of MZ twins and 559 pairs of DZ twins).\(^4\)

Descriptive statistics for the key variables employed in the analysis are presented in Table 1. Details on the construction of some of the main variables are contained in Appendix A. Computations of the statistics are made for the sample of twins, aged 20-64 years, who provide information on each of the variables used in the analysis. As presented in Table 1, the mean level of education of MZ twins is 12.51 years. Each twin tends to report a lower education level for his/her co-twin. The results show that the average education levels reported by co-twins are 12.37 years for MZ twins and 12.49 years for DZ twins.

As suggested by Table 1, there is a greater proportion of males in the sample of DZ twins.

\(^4\) Depending upon the specification of the model, the sample size can vary. That is, the inclusion of variables will result in a smaller sample size.
<table>
<thead>
<tr>
<th>Variable</th>
<th>MZ Twins (i)</th>
<th>DZ Twins (ii)</th>
<th>Total Sample (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported education (years)</td>
<td>12.514 (2.51)</td>
<td>12.720 (2.41)</td>
<td>12.612 (2.46)</td>
</tr>
<tr>
<td>Sibling-reported education (years)</td>
<td>12.371 (2.49)</td>
<td>12.489 (2.34)</td>
<td>12.428 (2.42)</td>
</tr>
<tr>
<td>Twins report same own level of education (proportion)</td>
<td>0.557 (0.50)</td>
<td>0.376 (0.48)</td>
<td>0.469 (0.50)</td>
</tr>
<tr>
<td>Report by co-twin same as self-report (proportion)</td>
<td>0.747 (0.43)</td>
<td>0.675 (0.47)</td>
<td>0.713 (0.45)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>36.832 (8.29)</td>
<td>35.317 (8.01)</td>
<td>36.099 (8.19)</td>
</tr>
<tr>
<td>Married (proportion)</td>
<td>0.760 (0.43)</td>
<td>0.724 (0.45)</td>
<td>0.742 (0.44)</td>
</tr>
<tr>
<td>Male (proportion)</td>
<td>0.468 (0.50)</td>
<td>0.504 (0.50)</td>
<td>0.486 (0.50)</td>
</tr>
<tr>
<td>Log of annual income</td>
<td>9.997 (0.30)</td>
<td>10.008 (0.30)</td>
<td>10.002 (0.30)</td>
</tr>
<tr>
<td>Sample size</td>
<td>1192</td>
<td>1118</td>
<td>2310</td>
</tr>
</tbody>
</table>

twins relative to MZ twins (50.4 percent compared with 46.8 percent). The mean ages for MZ and DZ twins are 36.83 years and 35.52 years, respectively. This indicates that DZ twins are, on average, 1.5 years younger than MZ twins. There is also a greater representation of MZ twins who are married (76.0 percent) relative to DZ twins (72.4 percent). The (logarithmic) annual income of DZ twins is quite similar to that of MZ twins, with both figures in Table 1 being around the value 10.
IV Empirical Results

Applying the appropriate estimation method to the models derived (either Ordinary Least Squares or Generalised Least Squares), the various results are presented in the following sub-sections.

Benchmark set of results

In this subsection, a conventional analysis of the twins data is presented. The conventional analysis is based on an estimating equation of the form:

\( y_i = \alpha_0 + \beta_1 \text{Educ}_i + \beta_2 \text{Gender}_i + \beta_3 \text{Age}_i + \beta_4 \text{Marital _ Status}_i + u_i. \)

This type of estimating equation is fairly common in the literature being the same as Miller et al. (1995) and broadly similar to Ashenfelter and Krueger (1994) and Taubman (1976). It is in the spirit of Mincer (1974), with the only departure from Mincer being the replacement of a quadratic in experience by a linear age term.\(^5\) As argued by Miller et al. (1995), this change is justified on empirical grounds.\(^6\)

To control for individual-specific environmental factors, drinking and smoking variables are included as proxy variables for the social behaviour of the respondents in equation (10). Thus, the estimating equation becomes

\(^5\) The so-called Mincer earnings equation is given as: \( y_i = \beta_0 + \beta_1 \text{Educ}_i + f(\text{Experience}_i) + \varepsilon_i \).

\(^6\) Given the method of constructing the dependent variable outlined in the Appendix, the linear age variable is an adequate representation of the growth in wages with time. On the question of the use of age rather than a labour market experience variable, Blinder’s (1976) argument can be used, namely that the choice of variables is an empirical matter. Here, because the data are pooled across males and females and the Mincer potential labour market experience variable is known to be a poor proxy for female labour market experience, it makes sense to use the age variable.
The drinking variable included in the estimating equation follows the specification adopted by Berger and Leigh (1988). Hence, using the information provided by the respondents on the average quantity of drinks consumed in a typical week, non-drinkers are defined as those who consume no alcohol at all.

A slightly different specification of the smoking effect is used in order to make greater use of the data available on smoking habits. These groups are distinguished. Non-smokers are defined as those who have never smoked cigarettes during their lifetime. Smokers refer to those who consume at least one cigarette, on average, a day. An additional category, the "ex-smoker" category, comprises respondents who claimed that they had stopped smoking at the time of the survey.

The first step in this analysis is to treat the entire sample of 1155 pairs of twins as a sample of individuals. An OLS regression is then performed on the sample of 2310 individuals. Column (i) of Table 2 lists the OLS estimates of the earnings function specified in equation (11). The OLS estimate of the return to schooling is 6.3 percent. The coefficient on the drinking dummy variable implies that drinkers earn 3.3 percent more than non-drinkers. The direction of this impact is consistent with the findings of Berger and Leigh (1988), though the magnitude of the wage effect is much lower than in that study. After correcting for selectivity bias, Berger and Leigh (1988) find that male drinkers earn 36.2 percent more than male non-drinkers while female drinkers have a 28.7 percent income advantage over female non-drinkers. The results are,

\[
\begin{align*}
\text{(11)} \quad y_i &= \alpha_0 + \beta_1 Edu_{ci} + \beta_2 Gender_i + \beta_3 Age_{ci} + \beta_4 Marital \_ Status_i + \\
&\quad \beta_5 Drink_i + \beta_6 Smoke_i + u_i.
\end{align*}
\]

The drinking variable included in the estimating equation follows the specification adopted by Berger and Leigh (1988). Hence, using the information provided by the respondents on the average quantity of drinks consumed in a typical week, non-drinkers are defined as those who consume no alcohol at all.

A slightly different specification of the smoking effect is used in order to make greater use of the data available on smoking habits. These groups are distinguished. Non-smokers are defined as those who have never smoked cigarettes during their lifetime. Smokers refer to those who consume at least one cigarette, on average, a day. An additional category, the "ex-smoker" category, comprises respondents who claimed that they had stopped smoking at the time of the survey.

The first step in this analysis is to treat the entire sample of 1155 pairs of twins as a sample of individuals. An OLS regression is then performed on the sample of 2310 individuals. Column (i) of Table 2 lists the OLS estimates of the earnings function specified in equation (11). The OLS estimate of the return to schooling is 6.3 percent. The coefficient on the drinking dummy variable implies that drinkers earn 3.3 percent more than non-drinkers. The direction of this impact is consistent with the findings of Berger and Leigh (1988), though the magnitude of the wage effect is much lower than in that study. After correcting for selectivity bias, Berger and Leigh (1988) find that male drinkers earn 36.2 percent more than male non-drinkers while female drinkers have a 28.7 percent income advantage over female non-drinkers. The results are,

\[
\begin{align*}
\text{(11)} \quad y_i &= \alpha_0 + \beta_1 Edu_{ci} + \beta_2 Gender_i + \beta_3 Age_{ci} + \beta_4 Marital \_ Status_i + \\
&\quad \beta_5 Drink_i + \beta_6 Smoke_i + u_i.
\end{align*}
\]

The drinking variable included in the estimating equation follows the specification adopted by Berger and Leigh (1988). Hence, using the information provided by the respondents on the average quantity of drinks consumed in a typical week, non-drinkers are defined as those who consume no alcohol at all.

A slightly different specification of the smoking effect is used in order to make greater use of the data available on smoking habits. These groups are distinguished. Non-smokers are defined as those who have never smoked cigarettes during their lifetime. Smokers refer to those who consume at least one cigarette, on average, a day. An additional category, the "ex-smoker" category, comprises respondents who claimed that they had stopped smoking at the time of the survey.

The first step in this analysis is to treat the entire sample of 1155 pairs of twins as a sample of individuals. An OLS regression is then performed on the sample of 2310 individuals. Column (i) of Table 2 lists the OLS estimates of the earnings function specified in equation (11). The OLS estimate of the return to schooling is 6.3 percent. The coefficient on the drinking dummy variable implies that drinkers earn 3.3 percent more than non-drinkers. The direction of this impact is consistent with the findings of Berger and Leigh (1988), though the magnitude of the wage effect is much lower than in that study. After correcting for selectivity bias, Berger and Leigh (1988) find that male drinkers earn 36.2 percent more than male non-drinkers while female drinkers have a 28.7 percent income advantage over female non-drinkers. The results are,

\[
\begin{align*}
\text{(11)} \quad y_i &= \alpha_0 + \beta_1 Edu_{ci} + \beta_2 Gender_i + \beta_3 Age_{ci} + \beta_4 Marital \_ Status_i + \\
&\quad \beta_5 Drink_i + \beta_6 Smoke_i + u_i.
\end{align*}
\]

The drinking variable included in the estimating equation follows the specification adopted by Berger and Leigh (1988). Hence, using the information provided by the respondents on the average quantity of drinks consumed in a typical week, non-drinkers are defined as those who consume no alcohol at all.

A slightly different specification of the smoking effect is used in order to make greater use of the data available on smoking habits. These groups are distinguished. Non-smokers are defined as those who have never smoked cigarettes during their lifetime. Smokers refer to those who consume at least one cigarette, on average, a day. An additional category, the "ex-smoker" category, comprises respondents who claimed that they had stopped smoking at the time of the survey.

The first step in this analysis is to treat the entire sample of 1155 pairs of twins as a sample of individuals. An OLS regression is then performed on the sample of 2310 individuals. Column (i) of Table 2 lists the OLS estimates of the earnings function specified in equation (11). The OLS estimate of the return to schooling is 6.3 percent. The coefficient on the drinking dummy variable implies that drinkers earn 3.3 percent more than non-drinkers. The direction of this impact is consistent with the findings of Berger and Leigh (1988), though the magnitude of the wage effect is much lower than in that study. After correcting for selectivity bias, Berger and Leigh (1988) find that male drinkers earn 36.2 percent more than male non-drinkers while female drinkers have a 28.7 percent income advantage over female non-drinkers. The results are,

\[
\begin{align*}
\text{(11)} \quad y_i &= \alpha_0 + \beta_1 Edu_{ci} + \beta_2 Gender_i + \beta_3 Age_{ci} + \beta_4 Marital \_ Status_i + \\
&\quad \beta_5 Drink_i + \beta_6 Smoke_i + u_i.
\end{align*}
\]

The drinking variable included in the estimating equation follows the specification adopted by Berger and Leigh (1988). Hence, using the information provided by the respondents on the average quantity of drinks consumed in a typical week, non-drinkers are defined as those who consume no alcohol at all.

A slightly different specification of the smoking effect is used in order to make greater use of the data available on smoking habits. These groups are distinguished. Non-smokers are defined as those who have never smoked cigarettes during their lifetime. Smokers refer to those who consume at least one cigarette, on average, a day. An additional category, the "ex-smoker" category, comprises respondents who claimed that they had stopped smoking at the time of the survey.

The first step in this analysis is to treat the entire sample of 1155 pairs of twins as a sample of individuals. An OLS regression is then performed on the sample of 2310 individuals. Column (i) of Table 2 lists the OLS estimates of the earnings function specified in equation (11). The OLS estimate of the return to schooling is 6.3 percent. The coefficient on the drinking dummy variable implies that drinkers earn 3.3 percent more than non-drinkers. The direction of this impact is consistent with the findings of Berger and Leigh (1988), though the magnitude of the wage effect is much lower than in that study. After correcting for selectivity bias, Berger and Leigh (1988) find that male drinkers earn 36.2 percent more than male non-drinkers while female drinkers have a 28.7 percent income advantage over female non-drinkers. The results are,
however, more in line with Hamilton and Hamilton (1997). They report that moderate drinkers earn 7 percent more than non-drinkers while heavy drinkers earn 14 percent more than non-drinkers.

Examination of the coefficients on the smoking dummy variables in column (ii) of Table 2 reveals that both current smokers and ex-smokers earn 3.3 percent and 2.6 percent less than non-smokers, respectively.

A broad summary of the results for the drinking and smoking variables is that they are of the same sign as findings reported in the overseas literature, but they are only around one-third to one-tenth the magnitude for the drinking variable and around one-half the magnitude for the smoking variables in the comparison studies. This may be a finding peculiar to the Australian labour market or it may be associated with the nature of the dependent variable. As noted earlier, the dependent variable in these analyses is the mean earnings of the worker's occupation of employment, and this implies that estimated coefficients will capture any inter-occupational earnings effects but not the intra-occupational earnings effects. While it is known from Groshen's (1991) work that a focus on the inter-occupational earnings differentials is appropriate when studying the returns to schooling, it is not known at this stage whether the wage effects of addictive behaviours are linked to inter-occupational or intra-occupational factors.  

Separating the samples into MZ and DZ twins, but still treating the two groups of twins as samples of individuals, an OLS regression is conducted. Column (ii) of Table 2 presents the OLS estimates for the 1192 MZ individuals. It is found that the

---

8 Research in progress using alternative data suggests that both sources of earnings differentials associated with addictive behaviours may be important.
Table 2. Ordinary Least Squares (OLS) Estimates of Log Annual Earnings for the Samples of 2310 Individuals, 1192 MZ Individuals and 1118 DZ Individuals: Australian Twin Survey.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Sample (i)</th>
<th>MZ Individuals (ii)</th>
<th>DZ Individuals (iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>9.214 (251.30)</td>
<td>9.237 (183.24)</td>
<td>9.182 (170.30)</td>
</tr>
<tr>
<td>Own education</td>
<td>0.063 (33.38)</td>
<td>0.063 (24.43)</td>
<td>0.064 (22.83)</td>
</tr>
<tr>
<td>Married</td>
<td>0.021 (1.98)</td>
<td>0.035 (2.38)</td>
<td>0.008 (0.51)</td>
</tr>
<tr>
<td>Age</td>
<td>0.002 (3.74)</td>
<td>0.002 (2.47)</td>
<td>0.002 (2.72)</td>
</tr>
<tr>
<td>Male</td>
<td>0.222 (24.36)</td>
<td>0.231 (18.31)</td>
<td>0.212 (15.87)</td>
</tr>
<tr>
<td>Drinking</td>
<td>0.033 (3.16)</td>
<td>0.015 (1.08)</td>
<td>0.054 (3.52)</td>
</tr>
<tr>
<td>Smoking – Current</td>
<td>-0.033 (2.78)</td>
<td>-0.034 (2.06)</td>
<td>-0.029 (1.65)</td>
</tr>
<tr>
<td>Smoking – Past</td>
<td>-0.026 (2.35)</td>
<td>-0.011 (0.71)</td>
<td>-0.043 (2.65)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.495</td>
<td>0.515</td>
<td>0.477</td>
</tr>
<tr>
<td>Sample size</td>
<td>2310</td>
<td>1192</td>
<td>1118</td>
</tr>
</tbody>
</table>

Numbers in parentheses are t statistics.

The coefficient on the schooling variable is 6.3 percent, the same order as for the total sample of individuals. The gender and age effects for MZ individuals are also similar in magnitude to those for the total sample, while the marital status effect is slightly stronger for MZ individuals than that for all individuals reported column (i).
An examination of the drinking and smoking variables indicates that only the variable indicating that the person is a current smoker is statistically significant, the estimated coefficient revealing that this group of individuals earn less than non-smokers by 3.4 percent. The insignificance of the variables for drinker and ex-smoker is an intriguing aspect of the results, and it will be discussed below.

For the sample of 1118 DZ individuals (see column (iii) of Table 2), the OLS estimate of the return to schooling is 6.4 percent. This 6.4 percent estimate of the return to schooling is similar to that estimated for the MZ sample. The results on age and gender are quite similar in the MZ and DZ samples. However, the coefficient on the marital status variable is statistically insignificant in the sample of DZ individuals.

As presented in column (iii) of Table 2, the three variables for drinking and smoking behaviour are all statistically significant at the 10 percent level or better. The point estimate for the variable reflecting drinking behaviour shows that alcohol consumption leads to an increase in earnings of 5.4 percent relative to abstention. The smoking dummy variables show that both current smokers and ex-smokers command lower earnings than individuals who do not smoke. Specifically, current smokers and ex-smokers earn 2.9 percent and 4.3 percent less than non-smokers, respectively. The estimated impact associated with current smoking behaviour is only at the margin of significance when a 10 percent level of significance is used, whereas that associated with past smoking behaviour is statistically significant at the 1 percent level.

The similarities and differences between the findings for MZ and DZ twins require further comments. In terms of similarities, the effects of educational attainment, age and gender are almost identical in the two samples. These are variables where there is reasonably strong agreement among economists on the rationale for inclusion in an earnings equation and on the empirical results. In terms of differences, the effects on
earnings of marital status differ appreciably between the samples of MZ and DZ twins. For this variable, there is not a consensus among economists on the rationale for inclusion in the Mincer-type earnings equation (see the rationales provided by Korenman and Neumark (1991)). Where marital status and the addictive behaviour variables are in fact proxy variables for a range of environmental and behavioural influences, the estimated impact might be expected to vary considerably across non-random samples. It is known that MZ and DZ twins differ with respect to various family characteristics. While MZ births occur randomly, DZ births vary by race, and age of the mother, among other factors (see Gedda (1961), pp. 68-77). For example, the incidence of DZ births is four times greater among mothers age 40 than for mother age 20 (Gedda (1961), p. 73). Consequently, DZ twins are more likely to be Catholic and have marriage patterns that may generate the different relationships between marital status and wages observed here. Many of these influences can be netted out in the fixed-effects model supplied in the next sub-section.

Fixed-Effects Analysis

Fixed-effects estimates for MZ and DZ twins are listed in columns (i) and (ii) of Table 3, respectively. The first important finding concerning the return to schooling is that the coefficient on the schooling variable for MZ twins is 0.024, which is statistically significant ($t = 4.79$). This estimate, which provides a measure of the returns to schooling net of genetic endowments and shared family environment, is about 40 percent of the conventional estimate (see column (ii) of Table 2).

The signs of the coefficients on the drinking and smoking dummy variables in the equation for MZ twins are consistent with the earlier findings. The coefficient on the

<table>
<thead>
<tr>
<th>Variable</th>
<th>MZ Twins (i)</th>
<th>DZ Twins (ii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own education</td>
<td>0.024</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(4.79)</td>
<td>(9.63)</td>
</tr>
<tr>
<td>Married</td>
<td>0.036</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>(1.83)</td>
<td>(0.71)</td>
</tr>
<tr>
<td>Age</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Male</td>
<td>-</td>
<td>0.222</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>(11.84)</td>
</tr>
<tr>
<td>Drinking</td>
<td>0.035</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>(1.81)</td>
<td>(2.79)</td>
</tr>
<tr>
<td>Smoking – Current</td>
<td>-0.049</td>
<td>-0.064</td>
</tr>
<tr>
<td></td>
<td>(1.88)</td>
<td>(2.48)</td>
</tr>
<tr>
<td>Smoking – Past</td>
<td>-0.002</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(2.43)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.058</td>
<td>0.363</td>
</tr>
<tr>
<td>Sample size</td>
<td>596</td>
<td>559</td>
</tr>
</tbody>
</table>

Numbers in parentheses are $t$ statistics.

(A) Variables not relevant.

past smoking variable is statistically insignificant, whereas the coefficients on the current smoking variable and the drinking variable are significant at the 5 percent level. While the estimated coefficient on the drinking variable for MZ twins is similar in the fixed-effects analysis as in column (i) of Table 2, the impact on earnings of being a current smoker in the fixed-effects model is around 50 percent greater (in absolute value) than in the results for individuals (see column (i) of Table 2). This
suggests that the significant wage effects recorded in equation estimated on samples of individuals (column (i) of Table 2) may compound the inter-relationships between addictive behaviour, genetic endowments and common family background rather than reflect the effects of smoking and drinking behaviour per se. These issues are explored further below.

For DZ twins (see column (ii)), the estimated return to schooling from the fixed-effects model is 4.4 percent. The difference of 1.9 percentage points between this estimate and the conventional estimate of 6.3 percent (column (i) of Table 2) gives an indirect measure of the bias in conventional measures of the return to schooling due to failure to control for shared family background. The effects of drinking and smoking behaviour on earnings relative to non-drinkers and non-smokers in the fixed-effects model for DZ twins are quite interesting.9 The variable for smoking reveals that current smokers are at a 6.4 percent income disadvantage compared with non-smokers. The variable for ex-smokers indicates that ex-smokers earn 5.5 percent less than non-smokers. In the case of drinking, the drinking dummy variable reveals that drinkers have a 6.1 percent income advantage compared with non-drinkers.

There are interesting patterns across the samples of individuals, DZ and MZ twins for the dummy variables recording the use of alcohol and cigarettes. In the case of current smoking, comparison of the relevant estimates in the three samples of individuals, DZ and MZ twins (see Table 4) shows that the reduction in earnings is greater for DZ twins (fixed-effects estimates) than for the sample treated as one of individuals (compare columns (i) and (ii)), while the impact smoking has on earnings

9 Using brothers, Levine et al. (1997) find that their fixed-effects estimates of the wage effects of smoking also increased. The comparability of the pattern across Levine et al.'s (1997) study and this study is very reassuring.
Table 4. Ordinary Least Squares (OLS) and Fixed-Effects Estimates of Log Annual Earnings: Australian Twin Survey.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS</th>
<th>Fixed-Effects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individuals</td>
<td>DZ Twins (i)</td>
<td>MZ Twins (ii)</td>
</tr>
<tr>
<td></td>
<td>(i)</td>
<td>(i)</td>
<td>(ii)</td>
</tr>
<tr>
<td>Drinking</td>
<td>0.033</td>
<td>0.061</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>(3.16)</td>
<td>(2.79)</td>
<td>(1.81)</td>
</tr>
<tr>
<td>Smoking – Current</td>
<td>-0.033</td>
<td>-0.064</td>
<td>-0.049</td>
</tr>
<tr>
<td></td>
<td>(2.78)</td>
<td>(2.48)</td>
<td>(1.88)</td>
</tr>
<tr>
<td>Sample size</td>
<td>2310</td>
<td>1192</td>
<td>1118</td>
</tr>
</tbody>
</table>

(a) columns (i), (ii), and (iii) present results from column (i) of Table 2, column (ii) of Table 3, and column (i) of Table 3, respectively.

(b) numbers in parentheses are t statistics.

of MZ twins is only significant when a 10 percent level of significance is used (see column (iii)). In the case of drinking, the point estimates are approximately the same for the sample treated as one of individuals and for MZ twins, and these estimates are around 2-3 percentage points less than those obtained from the fixed-effects model for DZ twins.

Comparing columns (i) and (ii) of Table 4, it is noted that the first set of estimates does not hold family background constant, whereas the adjacent set of estimates does. On the assumption that a favourable family background is associated with higher earnings, application of the standard omitted-variables formula suggests that family background (during childhood) is positively related to the measure of current smoking behaviour (where there is a 3.1 percentage point difference between the estimates in
columns (i) and (ii)) and negatively related to drinking behaviour (where the estimates in columns (i) and (ii) gives a negative 2.8 percentage point difference.\textsuperscript{10, 11}

The difference between columns (ii) and (iii) of Table 4 is that the set of estimates in column (iii) holds constant genetic endowments while the set of estimates in column (ii) does not. Both sets hold constant family background. The estimates in column (iii), relative to those in column (ii), are around 1.5 percentage points higher for the smoking variable, and around 2.6 percentage points lower for the drinking variable. Application of the standard omitted-variables formula suggests, therefore, that greater genetic endowments reduce the likelihood of smoking but increase the drinking behaviour analysed in Table 4 (on the assumption that genetic endowments are positively associated with earnings).\textsuperscript{12, 13}

Thus, these results suggest that genetics and family background affect earnings and also drinking and smoking behaviour. The positive effect on wages of drinking, and the negative effect of smoking reported in analyses of individuals in this study and in the literature are attributable to genetic endowments and family background as well as to drinking and smoking per se.

\textsuperscript{10} Adopting the suggestion by Agawal and Goedde (1990) of the possible influence of family background on drinking, it may be reasoned that the positive relationship between family background and current smoking behaviour arises because children try to model their behaviour on that of their parents and in doing so may also imitate their smoking habits.

\textsuperscript{11} Agawal and Goedde (1990) suggest that drinking may be discouraged in some families for religious, cultural, or climatic grounds, thus resulting in a negative relationship between family background and drinking behaviour.

\textsuperscript{12} Collins and Marks (1991) report that the reduction in the likelihood of smoking may be explained by genetic factors.

\textsuperscript{13} Partanen \textit{et al.} (1966) report that genetic endowments significantly increase the frequency and amount of drinking in their studies of twins between 28 and 37 years of age.
Selection-Effects Analysis

The selection-effects model is estimated in order to examine the robustness of the findings reported above and also because it provides an explicit treatment of the links between wages and family effects. As most of the findings parallel those reported above, the analysis will be brief.

Selection-effects estimates for the twins are presented in Table 5. Estimates for MZ twins are listed in column (i), and those for DZ twins are in column (ii). In the MZ sample, the coefficients on the own education and co-twin’s education variables are 0.047 and 0.023, respectively. The coefficient on the co-twin’s education variable is a measure of the selection effect. With a value of 0.023, the selection effect is positive and statistically significant. The meaning behind this positive value is that the better-educated families (in this data set) are those who would be the most highly compensated in the labour market. The positive selection effect may also mean that families with high earnings would be the most likely to educate their children. With these two coefficients, the selection-effects estimate of the return to schooling is computed as 0.024 (i.e., 0.047 - 0.023). This implies that any regression estimator that does not account for selection effect will be upward-biased. Note that this 0.024 estimate of the return to schooling is very similar to the fixed-effects estimate of the return to schooling of 0.024 reported in the previous section (see column (i) of Table 3).

Adjusting for selection effect results in an estimated wage effect of drinking of 0.037 \([= 0.019 - (-0.018)]\).\(^{14}\) As for smoking, the effect past smoking behaviour has

\(^{14}\) Each of the selection effects is insignificant. The point estimates are discussed here to facilitate comparison with the results from the fixed-effects model.
Table 5. Generalised Least Squares (GLS) Selection-Effects Estimates of Log Annual Earnings: Australian Twin Survey

<table>
<thead>
<tr>
<th>Variable</th>
<th>MZ Twins (i)</th>
<th>DZ Twins (ii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>9.121</td>
<td>9.042</td>
</tr>
<tr>
<td></td>
<td>(149.90)</td>
<td>(146.00)</td>
</tr>
<tr>
<td>Own education</td>
<td>0.047</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>(16.13)</td>
<td>(20.77)</td>
</tr>
<tr>
<td>Co-twin's education</td>
<td>0.023</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(7.96)</td>
<td>(5.19)</td>
</tr>
<tr>
<td>Married</td>
<td>0.040</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(2.86)</td>
<td>(0.58)</td>
</tr>
<tr>
<td>Age</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(2.73)</td>
<td>(3.17)</td>
</tr>
<tr>
<td>Male</td>
<td>0.226</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>(15.63)</td>
<td>(16.58)</td>
</tr>
<tr>
<td>Own drinking</td>
<td>0.019</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>(1.40)</td>
<td>(3.53)</td>
</tr>
<tr>
<td>Co-twin's drinking</td>
<td>-0.018</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(1.32)</td>
<td>(0.52)</td>
</tr>
<tr>
<td>Own smoking – Current</td>
<td>-0.035</td>
<td>-0.034</td>
</tr>
<tr>
<td></td>
<td>(2.11)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>Co-twin’s smoking – Current</td>
<td>0.015</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>(0.88)</td>
<td>(1.85)</td>
</tr>
<tr>
<td>Own smoking – Past</td>
<td>-0.010</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>(0.62)</td>
<td>(2.71)</td>
</tr>
<tr>
<td>Co-twin’s smoking – Past</td>
<td>-0.006</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td>(0.73)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.534</td>
<td>0.490</td>
</tr>
<tr>
<td>Sample size</td>
<td>1192</td>
<td>1118</td>
</tr>
</tbody>
</table>

Numbers in parentheses are $t$ statistics.
on wages is \(-0.004 \approx -0.010 - (-0.006)\). Current smoking behaviour results in a decrease in wages by 5 percent [i.e., \((-0.035 - 0.015) \times 100\%\)]. A close examination of the drinking and smoking variables in the selection-effects model reveals that the estimates are very similar to those in the fixed-effects model for MZ twins (see column (i) of Table 3).

The coefficients on the other variables, as listed in column (i) of Table 5, are all statistically significant and of the expected sign. The results are also quite similar to those of the ordinary least squares and fixed-effects estimations for the MZ sample.

Column (ii) of Table 5 lists the estimates of the selection-effects model for DZ twins. The selection effect, which is statistically significant, is 0.015. The coefficient on the own education variable is 0.060. Adjusting for selection effect, the estimated return to schooling is 0.044 (i.e., 0.059 - 0.015). Similar to the MZ case, the selection-effects estimate of the return to schooling is very similar to the fixed-effects estimate of the return to schooling for the DZ sample (compare with column (ii) of Table 3). Note also that the positive selection effect suggests that any regression estimator that neglects the selection effect will be upward-biased.

The adjustment for selection effect in the DZ sample leads to drinkers having a 6.1 percent [i.e., \((0.053 - (-0.008)) \times 100\%\)] income advantage compared with abstainers. Workers who are ex-smokers earn 5.5 percent [i.e., \((-0.043 - 0.012) \times 100\%\)] less than non-smokers. Current smoking behaviour leads to a decrease in earnings of 6.6 percent relative to non-smoking. Again, the results obtained do not differ greatly from those obtained from the fixed-effects estimation for DZ twins (see column (ii) of Table 3).

These results for the wage effects of drinking and smoking, which are consistent with the literature, suggest that smoking has a deleterious effect on earnings but that
drinking is associated with higher earnings. A close examination of the two sets of selection-effects estimates of the drinking and smoking variables for MZ and DZ twins (see columns (i) and (ii) of Table 5) indicate the same patterns as those of the fixed-effects estimates (as presented in Table 4). The addictive behaviour results in Table 5 suggest that family background is positively related to current smoking behaviour and negatively related to drinking behaviour (assuming that a favourable family background is associated with higher earnings). It is suggested that genetic endowments increase the likelihood of drinking but reduce the incidence of smoking.

V Summary and Conclusion

The employment of the fixed-effects or selection-effects models reduces the estimated return to schooling from around 6.0 percent (OLS estimates) to around 2.4 percent in the case of MZ twins, and to around 4.4 percent in the case of DZ twins. These differences, which are substantial, show the importance of both genetic endowments and shared family environment in determining earnings, and the extent of bias in the return to schooling when both factors are not controlled for. They are suggestive of the need to also control for the influence of genetic endowments and shared family environment when studying the wage effects associated with other

---

15 In the MZ sample, the coefficients on the co-twin's drinking and smoking variables are statistically insignificant (see column (i) of Table 5). The insignificant coefficients indicate that the wage effects of drinking and smoking will not be biased downward if selection effects are not taken account of. For DZ individuals, the coefficient on the co-twin's drinking variable is statistically insignificant, but that on the smoking variable is significant at the 10 percent level. Therefore, failure to adjust for selection effects will result in an upward bias in the estimated wage effect of smoking but will not cause any bias in the coefficient on the drinking variable.
individual characteristics, including the addictive behaviours that are the focus of the current study.

The augmentation of the estimating equation with the respondents’ drinking and smoking behaviour reveals that drinking leads to an increase in earnings relative to abstention. This is consistent with the findings of Berger and Leigh (1988) and Hamilton and Hamilton (1997). Both past and present smoking habits have negative impacts on earnings. Levine et al. (1997) also report a negative wage effect of smoking.

On the assumption that a favourable family background is associated with higher earnings, application of the standard omitted-variables formula to the results obtained suggests that family background (during childhood) is positively related to the measure of current smoking behaviour and negatively related to drinking behaviour. Application of the standard omitted-variable formula also suggests that greater genetic endowments reduce the likelihood of smoking, but increase the drinking behaviour (on the assumption that there is a positive association between genetic endowments and earnings).

Thus, the effects of drinking and smoking on the occupational mean wages analysed in this study are attributable to genetic endowments and family background as well as to drinking and smoking per se. The effects in this regard are less pronounced than is the case for the impact of schooling. They are, however, sufficiently important that account should be taken of them where possible.
APPENDIX A  
Construction of Variables

The construction of some of the key variables is provided below.

Educational Attainment
The education data were gathered in categorical form with seven different categories (<7 years of schooling; 8-10 years of schooling; 11-12 years of schooling; apprenticeship, diploma, certificate; technical or teachers college; university, first degree; university, postgraduate degree). The present analysis uses education variables that were recoded as 5, 9, 11.5, 11.5, 13, 15, and 17 years of education, respectively. Miller et al. (1995) point out that the recoding of the education variables affects a number of individual findings, such as reliability ratios. However, the general gist of the conclusions derived from the study is not severely affected by reasonable variations in the assumed mean levels of education for each category.

Income
Due to the deficiency in the income data described in the text, an alternative income variable is used. This is a grouped income variable that was constructed using the average earnings of the occupation in which the individual was employed. The average income of full-time workers in each of the 60 minor group occupations was obtained from the 1986 Census of Population and Housing. This measure was applied to all members of the sample, regardless of whether they were employed on a full-time or part-time basis. This procedure is argued by Miller et al. (1995) to control for hours of leisure, and prevents estimates of returns to education being distorted by the presence of individuals who work more than the standard number of hours.

Other studies that have used the average earnings of the occupation as the dependent variable (or grouped income variable) include Griliches (1977), Nickell (1982), and Behrman et al. (1994). The grouped income variable captures the inter-occupational, but not the intra-occupational earnings effects. According to Groshen (1991), most of the impact schooling has on earnings comes about through inter-occupational factors. The relevant literature (e.g., Miller et al. (1995)) points out that Groshen's finding suggests that the method used in the construction of the dependent variable is only a minor limitation that is unlikely to affect the analysis of the twins sample treated as a sample of individuals, and results reported from study of earnings based on data sets that contain information on the individual's earnings (e.g., Preston (1997)) supports this stance. Whilst not ideal, it is a practical alternative that permits a thorough analysis of the Australian Twin Registry data. It is, however, a feature that needs to be kept in mind when interpreting some statistics. For example, because the within-occupation variation is averaged out of the income data, measures of goodness of fit in the regression analysis are much higher than is usually reported.
REFERENCES


28